

UNIVERSIDAD CARLOS III DE MADRID
Departamento de Biblioteconomía y Documentación
Programa de Doctorado en Documentación



**METADATOS Y RECUPERACIÓN DE
INFORMACIÓN: ESTÁNDARES,
PROBLEMAS Y APLICABILIDAD EN
BIBLIOTECAS DIGITALES**

Tesis doctoral presentada por:
Eva M^a Méndez Rodríguez

Director: Prof. Dr. D. José Antonio Moreiro González

Getafe (Madrid). Diciembre 2001

**METADATOS Y RECUPERACIÓN DE INFORMACIÓN:
ESTÁNDARES, PROBLEMAS Y APLICABILIDAD EN
BIBLIOTECAS DIGITALES**

A mis padres

*I saw that one enquiry only gave occasion to another,
that book referred to book,
that to search was not always to find,
and to find was not always to be informed.*

Samuel Johnson (1709-1784), 1753

*Knowledge of words, and ignorance of the Word.
All our knowledge brings us nearer to death,
But nearness to death no nearer to God.
Where is the Life we have lost in living?
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?*

Thomas Stearns Eliot (1888-1965). *The Rock* (1934)

AGRADECIMIENTOS

Antes de adentrarme en el desarrollo de esta investigación, quiero agradecer vivamente a quienes me han apoyado y ayudado en el transcurso de la misma. En primer lugar a mi director, amigo y maestro el **Dr. José Antonio Moreira**, por su consejo y respaldo, quien ha confiado siempre en mí, no sólo en la realización de esta tesis doctoral, sino a través de toda mi trayectoria académica y profesional. A la **Dra. Mercedes Caridad**, por ser siempre el timón y el aliento no sólo de mi investigación sino de todas las que se han realizado en el seno de este Departamento; a **Carmen López de Sosoaga** por su amistad, por las deliciosas horas de conversación y por cargar de credibilidad mis proyectos; a la **Dra. Murtha Baca** del *Getty Information Institute* por su generosidad y por ofrecerme su experiencia, su afecto y su apoyo en la distancia; al **Dr. José Carlos Rovira**, vicerrector de la Universidad de Alicante y director de la Biblioteca Virtual *Miguel de Cervantes* durante el período de redacción de este trabajo, por todas las facilidades que me ha dado y la confianza que ha depositado en mí para plantear un proyecto de aplicación de metadatos en esta biblioteca digital; asimismo quiero agradecer al maravilloso **equipo de Cervantes Virtual**, en especial a **Alex Bia**, subdirector de informática y responsable de innovación, por su interés, por su tiempo y por hacerme comprender que la colaboración entre bibliotecarios e informáticos no sólo es necesaria, también es posible.

Al **Departamento de Biblioteconomía y Documentación**, particularmente al director, el Dr. Elías Sanz y al director del doctorado, el Dr. Miguel A. Marzal, porque ha sido el entorno idóneo para alentar y legitimar una investigación de estas características. A todos mis compañeros, especialmente a **Tony, Tomás, Boni, Menchu** y **David**, con quienes comparto inquietudes y proyectos, por prestarme todo el apoyo y su afecto sincero, en los largos días que ha durado este trabajo.

Quiero agradecer también, muy especialmente y a nivel personal, a **Liliana Maspóns**, a **Laura López** y a **Coro Pichardo** por su paciencia, cariño y compañía en los momentos más difíciles y por jugar en mi equipo en la dura partida de la vida. A todos **mis amigos** (Richard, M^a José, Ángela Sorli, José Merlo, César, Ainhoa, Iñigo, Sandra, Belén, Adela, Carolina Rojas, Emma, Jesús, Patricia, Zoila, Ana M^a, etc.) que siempre han creído en mí y con los que he podido contar en todo momento, en las adversidades que surgen en un trabajo tan dilatado. Finalmente quiero agradecer a todos aquellos que, sin entender muy bien lo que hacía, como **mis padres** a quienes dedico esta tesis, han sufrido mis ausencias para que esta investigación se materializara en lo que a continuación se expone.

ÍNDICE GENERAL

Índice de tablas	p.XVII
Índice de gráficos	p.XVIII
Índice de figuras.....	p.XIX

1. Introducción	p.1
2. Objetivo, método y fuentes utilizadas	p.17
2.1. Marco y objetivos del trabajo.....	p.17
2.2. Aspectos metodológicos.....	p.26
2.3. Fuentes y bibliografía.....	p.43

PARTE I: De la información a la metainformación: metadatos

3. Metadatos: concepto, fundamento, aplicación y tipología	p.61
3.1. Precisiones terminológicas y uso histórico del término	p.61
3.2. Concepto	p.69
3.3. Fundamento de la metainformación: el Documento como Objeto	p.92
3.4. Aplicación de los metadatos.....	p.100
3.5. Tipos de metadatos.....	p.114
4. El problema de la asignación de metadatos.....	p.143
4.1. De la catalogación a la asignación de metadatos: reflexiones en torno al proceso técnico de los recursos electrónicos.....	p.146
4.2. Quién crea los metadatos	p.169
4.3. Cómo asociar metadatos a los recursos de información electrónica: herramientas y aplicaciones.....	p.196
4.3.1. Herramientas destinadas a la creación de metaetiquetas	p.199
4.3.1.1. Plantillas	p.200
4.3.1.2. Software cliente.....	p.204
4.3.2. Herramientas destinadas a la creación de metadatos	p.226
4.3.2.1. Plantillas	p.228
4.3.2.2. Software cliente.....	p.244
4.3.2.3. Creación de metadatos en SIG-DLOs	p.269
4.3.2.4. Otras aplicaciones de creación de metadatos	p.279
4.3.3. Consideraciones para la elección de un software.....	p.289

PARTE II: Normalización y metadatos: formatos, esquemas o estándares

5. Modelos de metadatos	p.297
5.1. Lenguajes de marcas y estructuración de la información	p.300
5.1.1. SGML (<i>Standard Generalized Markup Language</i>).....	p.306
5.1.2. De HTML a XML.....	p.310
5.1.3. El intercambio de información de recursos electrónicos ampliamente estructurados	p.322
5.2. Propuestas e iniciativas de Metadatos	p.328
5.2.1. Metadatos de propósito general	p.330
5.2.1.1. Estructuras genéricas de metainformación	p.331
5.2.1.2. DCMI (<i>Dublin Core Metadata Initiative</i>).....	p.340
5.2.2. Metadatos de propósito específico.....	p.359
5.2.2.1. TEI Header (<i>Text Encoding Initiative</i>)	p.360
5.2.2.2. Otros esquemas de metadatos	p.377
5.2.3. RDF: un «metamodelo» de metadatos.....	p.384
6. Normalización	p.403
6.1. Metadatos y el problema de la estandarización	p.405
6.2. Estándares de metadatos: casuística y formalización	p.422
6.3. Desarrollo de un estándar formal: el caso del DCMES	p.440
6.4. Aproximación a la normalización del vocabulario	p.449

PARTE III: Acceso global a la información: recuperación, metadatos y bibliotecas digitales

7. Metadatos y recuperación de información en Internet.....	p.471
7.1. Recuperación de información en la Red: consideraciones generales	p.474
7.1.1. <i>Browsing</i> y navegación: aspectos distintivos	p.479
7.1.2. Paradigma de interrogación: el problema de la relevancia	p.484
7.2. Tendencias de los sistemas de búsqueda Web... ¿metadatos?	p.490
7.3. Búsqueda en texto completo <i>versus</i> búsqueda en metadatos: ventajas e inconvenientes	p.518
8. Bibliotecas digitales y metadatos.....	p.535
8.1. ¿Biblioteca digital/virtual? La biblioteca como concepto y la diversidad designativa	p.539
8.2. Bibliotecas virtuales: del oxímoron del término a la realidad práctica.....	p.563
8.2.1. Bibliotecas productoras/poseedoras de su colección digital.....	p.568
8.2.2. <i>Subject gateways</i> y el concepto de colección distribuida	p.578
8.3. Bibliotecas digitales y globalización	p.590
8.3.1. El acceso universal a la información.....	p.592
8.3.2. Metadatos e interoperabilidad: la necesidad de integración	p.597

**PARTE IV: Proyecto de aplicación de metadatos en bibliotecas digitales:
Gestión de una colección digital a través de un sistema descriptivo basado
en metadatos**

BIBLIOTECA VIRTUAL

9. Diseño de un sistema de metadatos para la	
Biblioteca Virtual Miguel de Cervantes	p.621
9.1. Consideraciones previas	p.622
9.2. Contexto institucional y técnico del proyecto.	p.627
9.2.1. Marco del proyecto: importancia de la BVC (1999-2001)	p.628
9.2.2. La BVC como creadora y gestora de información digital.....	p.636
9.3. Propuesta de aplicación de metadatos en la BVC	p.648
9.3.1. Justificación y objetivos	p.649
9.3.2. Estado actual de la metainformación	p.656
9.3.3. Valoración y selección de un modelo de metadatos	p.673
9.3.4. Prospectiva y viabilidad del proyecto	p.684
10. Conclusiones	p.689
Bibliografía	p.705
Glosario explicativo de la terminología utilizada	p.771
Índice de siglas y acrónimos	p.815

ÍNDICE DE TABLAS

Tabla 1. Bases de datos: resultados en todos los campos	p.46
Tabla 2. Publicaciones en el 2001 (Inspec-Abril / Lisa-Junio).....	p.48
Tabla 3. Bases de datos: resultados en el campo de título (TI).....	p.49
Tabla 4. Bases de datos: Resultados en el campo de resumen (AB).....	p.50
Tabla 5. Bibliografía en la Web (en inglés).....	p.52
Tabla 6. Bibliografía en la Web (en español)	p.53
Tabla 7. Doble perspectiva del papel de los metadatos	p.100
Tabla 8. Tipos de metadatos en bandas (Dempsey y Heery).	p.136
Tabla 9. Interpretación del <i>continuum</i> de metadatos (Sutton)	p.137
Tabla 10. Estructura de directorios e información de TagGen Office	p.250
Tabla 11. Estructuras de datos y metadatos	p.330
Tabla 12. Elementos del modelo Dublin Core.....	p.345
Tabla 13. Interpretación de los elementos principales de la TEIH en XML.....	p.370
Tabla 14. Características y tendencias de los SRII mencionados	p.515
Tabla 15. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en inglés (Altaviasta y Google).....	p.555
Tabla 16. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en español (Altaviasta y Google)	p.556
Tabla 17. Resumen DAFO de la BVC ante la aplicación de metadatos	p.686

ÍNDICE DE GRÁFICOS

Graf. 1. Aumento de las publicaciones sobre "metadatos" en LISA e Inspec (1980-2001)	p.47
Graf. 2. Aumento de las publicaciones sobre "bibliotecas digitales" en LISA e Inspec (1980-2001)	p.48
Graf. 3. Aumento de las publicaciones sobre "metadatos" en el campo título de LISA e Inspec (1980-2001).....	p.50
Graf. 4. Aumento de las publicaciones sobre "bibliotecas digitales" en el campo título de LISA e Inspec (1980-2001).....	p.50
Graf. 5. Aumento de las publicaciones sobre "metadatos" en el campo resumen de LISA e Inspec (1980-2001).....	p.50
Graf. 6. Aumento de las publicaciones sobre "bibliotecas digitales" en el campo resumen de LISA e Inspec (1980-2001)	p.50
Graf. 7. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en inglés (Altavista).....	p.555
Graf. 8. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en inglés (Google).....	p.555
Graf. 9. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en español (Altavista).	p.556
Graf. 10. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en español (Google).....	p.556
Graf. 11. Formas de aparición del anglicismo WWW	p.810

ÍNDICE DE FIGURAS

Fig. 1. Triple visión de la aplicación de los metadatos.....	p.114
Fig. 2. CORC. Visualización DC de un registro, exportable a RDF.....	p.167
Fig. 3. Creación automática de metadatos con Klarity.....	p.179
Fig. 4. Free MetaTag Builder.....	p.201
Fig. 5. Free MetaTag Builder (resultados).....	p.201
Fig. 6. MetaTag Builder-Position-it.....	p.201
Fig. 7. MetaTag Builder-Position-it (resultados).....	p.201
Fig. 8. Free MetaTag Builder. Submitmaster (edición y resultados).....	p.202
Fig. 9. MetaManager. Site Owner.....	p.202
Fig. 10. MetaManager. Site Owner (resultados).....	p.202
Fig. 11. Meta Builder 2.....	p.202
Fig. 12. Meta Builder 2 (resultados).....	p.202
Fig. 13. 9X MetaMagic 2.07.....	p.205
Fig. 14. Auto Meta Tags (automático).....	p.210
Fig. 15. Auto Meta Tags (manual).....	p.210
Fig. 16. MetaDemon 2.1.1.....	p.212
Fig. 17. Color and Metatag Buster 2.02.....	p.215
Fig. 18. Meta Builder 6.1.2.....	p.217
Fig. 19. MetaFormer 2.0.0.....	p.219
Fig. 20. Promotion-Pro 4.6 (pantalla principal).....	p.221
Fig. 21. Promotion-Pro 4.6 (<i>keywords</i>).....	p.223
Fig. 22. Meta Tag Maximizer 2.1.....	p.226
Fig. 23. Nordic Metadata Project Template.....	p.230
Fig. 24. Nordic Metadata Project (resultados).....	p.230
Fig. 25. Procesos de trabajo con Reggie Metadata Editor.....	p.232
Fig. 26. Reg (opciones).....	p.235
Fig. 27. Reg (plantilla según DC).....	p.235

Fig. 28. DC-Dot. Metadatos extraídos automáticamente y plantilla de edición...	p.239
Fig. 29. DC-assist v.1.6. (Elemento <i>coverage</i>).....	p.239
Fig. 30. RSS-Express. Canal Xml.com	p.241
Fig. 31. BetaMeta Bureau.....	p.242
Fig. 32. Creación de metadatos con NoteTab.....	p.244
Fig. 33. Hisoftware File List Builder.....	p.249
Fig. 34. Hisoftware Profile Builder	p.249
Fig. 35. TagGen 4.6 Professional	p.251
Fig. 36. TagGen Dublin Core Editor	p.253
Fig. 37. TagGen Dublin Core Editor (conversión MARC).....	p.253
Fig. 38. Metacontents™ Builder	p.254
Fig. 39. Hi-Verify™ 3.0	p.257
Fig. 40. JITI Builder® for Microsoft Office 2000 & 97	p.258
Fig. 41. Metabrowser.....	p.260
Fig. 42. Metabrowser (Tree Viewer – DC Schema).....	p.262
Fig. 43. Metabot Professional	p.264
Fig. 44. Metabot Professional. Selección de etiquetas para informe.....	p.266
Fig. 45. Metamanage	p.268
Fig. 46. MetaStar Data Entry. Creación de una plantilla de metadatos	p.273
Fig. 47. Esquema genérico de los módulos de un SIG-DLOs	p.279
Fig. 48. Metalite (español)	p.283
Fig. 49. GEMCat (Thesaurus)	p.285
Fig. 50. MARCit. Opciones de configuración.....	p.286
Fig. 51. d2m (Dublin Core to MARC converter).....	p.288
Fig. 52. d2m. Resultados en formato TXT.....	p.288
Fig. 53. Expresión gráfica del modelo de datos RDF.....	p.389
Fig. 54. Protégé 2000	p.399
Fig. 55. Oingo (resultados de búsqueda).....	p.505
Fig. 56. Interfaz de Hotmeta (1998)	p.528

Fig. 57. Interfaz de Hotmeta (1999-2001, v.1.4.6)	p.528
Fig. 58. Visión de la biblioteca digital según las proposiciones clásicas de la RI	p.533
Fig. 59. Ciberoteca. Bancaja (pantalla principal)	p.552
Fig. 60. Reformulación de la biblioteca como sistema en el mundo digital (Bibliotecas productoras-poseedoras de su colección digital)	p.572
Fig. 61. Reformulación de la biblioteca como sistema en el mundo digital (Colecciones virtuales/ <i>Subject gateways</i>).....	p.582
Fig. 62. Expresión gráfica del concepto de <i>biblioteca virtual</i> aplicado a la BVC	p.636
Fig. 63. <i>Workflow</i> de producción de la BVC.....	p.641
Fig. 64. Ficha de catálogo de obras de la BVC: <i>Las Esmeraldas</i> de J. Dicenta (Web)	p.644
Fig. 65. Recuperación de información y catálogo en la BVC.....	p.647
Fig. 66. Catálogo de títulos en la BVC	p.647
Fig. 67. Propuesta de visualización de metadatos de usuario en la BVC.....	p.682
Fig. 68. Propuesta de <i>Workflow</i> de la BVC incluyendo la creación de metadatos	p.683
Fig. 69. HCL (Metabrowser)	p.784

CAPÍTULO 1:

INTRODUCCIÓN

La Sociedad de la Información y la globalización son inminentes en el siglo XXI. La Biblioteconomía está en ebullición gracias a la aceptación y adopción de la tecnología. Todos los problemas, desafíos y sinergias que esto conlleva no son accidentales, son el resultado del rápido desarrollo del mundo de la información. Estas afirmaciones están repletas de retos y oportunidades para el profesional bibliotecario. De igual forma que el fin del siglo XIX y el principio del XX comportaron el mayor cambio de paradigma en las clasificaciones documentales, en el nuevo milenio, la información electrónica presenta una arquitectura hipertextual, no lineal y distribuida que trae consigo un nuevo cambio en el prototipo de la organización del conocimiento.

Por todo ello, en este trabajo de investigación abordamos el tema que está explícito en el título del mismo: los *metadatos y la recuperación de información*, analizando todos los problemas que comporta, tanto a nivel conceptual como de desarrollo de nuevas formas de estructurar la información, centrándonos para ello en la cuestión fundamental de la normalización y en las posibilidades de aplicación de metainformación para organizar las nuevas bibliotecas digitales. Los **metadatos**, la necesidad de etiquetar, *catalogar* y describir información estructurada, de tal forma que permita que los objetos Web se puedan almacenar, localizar, procesar e intercambiar a través del ordenador, protagonizarán sin duda, un nuevo paradigma en los sistemas de información del milenio recién estrenado. Este tema ha despertado un gran interés y debate científico en la comunidad bibliotecaria internacional y en todo el sector emergente de los sistemas de información electrónica debido a la evidencia

de Internet como fuente, depósito y recurso de información, así como a las nuevas formas de registrar el conocimiento que conllevan la perentoria explosión de la información digital, cuya idiosincrasia difiere tanto de la documentación impresa. Todos estos cambios han puesto en tela de juicio los métodos tradicionales de procesamiento de la información (fundamentalmente, catalogación, clasificación e indización), y el propio papel del bibliotecario en dicho proceso. Estas transformaciones están haciendo que los profesionales de la información reconozcan que sólo a través de una nueva definición estratégica de las funciones bibliotecarias esenciales (selección, organización y acceso) y de la propuesta de un nuevo modelo estructural, cuyo pilar, desde nuestro punto de vista, son los metadatos, se podrá afrontar el cambio de una manera eficiente. En esta tesis doctoral se refleja toda esta dialéctica, constituyendo una aportación en el ámbito de la necesidad y de la aplicabilidad de los metadatos en bibliotecas digitales.

Este planteamiento, que parece tan categórico, comporta sin embargo múltiples dificultades, hipótesis y preguntas que intentamos recoger y responder en esta investigación y que resumimos a continuación:

- En primer lugar, estamos aludiendo a un cambio de paradigma en la concepción de sistemas de información. Todo cambio supone un esfuerzo, y ese esfuerzo, en lo que a organización de la información se refiere, consistirá en una predefinición de conceptos y de tareas, a la par que un esfuerzo en la asunción de un nuevo papel del bibliotecario en la gestión de la información. En este sentido surgen preguntas como ¿qué son realmente los metadatos?, ¿sobre qué operan, cuál es su fundamento?, ¿para qué sirven?, ¿cómo se trabaja con metadatos?, ¿qué herramientas y destrezas se necesitan para procesar la metainformación?
- En segundo lugar, el problema de la normalización se ha contemplado en el ámbito de la Documentación desde siempre, ocupando a profesionales de la información durante décadas en la tarea facilitar el intercambio y recuperación de información. A medida que la tecnología Web se va adoptando, también cabe preocuparse por los mecanismos de estandarización que serán necesarios para hacer que esa información estructurada y de calidad sea visible en Internet gracias a los metadatos. Las preguntas relativas a la normalización no son menos profundas ¿Qué formato estandarizado deben tener los metadatos? ¿Qué nivel de normalización deben tener las nuevas estructuras de metainformación digital? ¿Todos los modelos de metadatos circunscritos a un dominio de aplicación deberán basarse en el mismo lenguaje o metalenguaje para permitir la interoperabilidad en la nueva Sociedad de la Información Global?
- Por otra parte, el problema de la recuperación de información es inmanente a nuestra disciplina y promotor de investigación. Teniendo en cuenta la amplitud y complejidad que conlleva este aspecto, tratamos de circunscribir este problema al contexto de las bibliotecas digitales, sin que ello obste para hacer un planteamiento genérico de la recuperación de información en Internet, donde los metadatos ocuparon, en su primera etapa de desarrollo (1994-1998), el papel de

“promesa” para la promoción y recuperación genérica en la WWW y cobran una nueva dimensión ante la proyección de la Web semántica.

- Por último, la propia noción de *cambio de paradigma* conlleva un conjunto coherente y relativamente autónomo de premisas referentes a modelos —en este caso de organización y recuperación de información— que proporcionan un alegato a diferentes teorías o a marcos teóricos. Lo propio del paradigma y del cambio en éste, será su carácter fundamentante, su condición de modelo organizador y suministrador de una orientación epistemológica. Si la organización de la información se ha considerado tradicionalmente una de las bases epistemológicas de la Documentación¹, los metadatos, como estructuras para organizar y recuperar la información digital, serán el elemento protagónico de este cambio de paradigma.

La evolución del procesamiento de la información y la profunda integración de las estructuras sociales y los sistemas de información están aumentando la importancia de la información y del conocimiento en un nivel muy alto. El acceso a la información por cualquier ciudadano es la esencia de los servicios de información de carácter universal y el objetivo básico de las políticas de información actuales. Los países democráticos tienden a dirigir sus sistemas de información hacia una Infraestructura Global de Información, hacia una **Sociedad de la Información Global**. Este hecho promueve inquietudes, tendencias y un sinfín de perspectivas de futuro, sobre todo en el ámbito de los estándares y de la interoperabilidad entre sistemas informativos.

Si Ortega y Gasset² decía que el concepto de biblioteca moderna (en 1935) surgió en torno a 1850 cuando la sociedad satisface la necesidad de contar con libros

¹ Rafael Capurro. *Epistemology and Information Science*. REPORT TRITA-LIB-6023. Stephan Schwarz, ed. August 1985 [documento HTML]. Disponible en: <http://www.capurro.de/trita.htm> (consultado el 15 de septiembre de 1999).

² José Ortega y Gasset. *La misión del bibliotecario*. Málaga: Asociación Andaluza de Bibliotecarios, 1994, p. 22 y ss.

y se genera la necesidad de buscar lectores. De forma análoga, podemos decir que hoy, a principios del siglo XXI, en concepto de biblioteca moderna está estrechamente ligado al concepto de biblioteca digital y surge cuando la sociedad satisface su necesidad de contar con información, incluso sufriendo una inflación informativa, y se genera la necesidad de buscar conocimiento. En ambos casos, el Estado asume la responsabilidad de satisfacer las necesidades sociales, entonces, a mediados del siglo XIX, creando bibliotecas para los lectores, y ahora, en el comienzo del siglo XXI, facilitando el acceso a la información, al patrimonio cultural y, en definitiva, al conocimiento a todos los ciudadanos a través de las infraestructuras de telecomunicaciones.

Hace ya algunas décadas, el desarrollo de los ordenadores y de las tecnologías de las comunicaciones crearon las condiciones apropiadas, la infraestructura idónea, para el procesamiento técnico de la información. La principal ventaja de ese procesamiento técnico fue una potencia mucho mayor en términos de velocidad y de cantidad de datos tratados. En los últimos años estamos asistiendo a una simbiosis entre el procesamiento técnico y el procesamiento social de la información, lo que conlleva la formación de complejos sistemas socio-técnicos de procesamiento de información que combinan la inteligencia humana, con complicadas aproximaciones al problema de la recuperación de datos. Conducir la práctica bibliotecaria a este nuevo entorno social protagonizado por la digitalización de la información y/o la producción de información digital es un experimento insoslayable y una nueva hipótesis de trabajo.

La información y las tecnologías aplicadas a la información se están convirtiendo, se han convertido ya, en un punto crítico para la ciencia, la investigación y la empresa de nuestro tiempo. Existe una tendencia incipiente hacia el acceso universal a las fuentes de información posibilitada por el crecimiento y desarrollo de las redes de telecomunicaciones, especialmente de Internet y, sobre todo, de su colección de recursos multimedia conocida como World-Wide Web. En menos de diez años, la WWW ha pasado de ser un sistema casi esotérico, utilizado por una comunidad limitada de investigadores estadounidenses y europeos, a ser un

almacén de datos ingente, universal, abierto, multitemático y multilingüe, así como a considerarse una de las herramientas por antonomasia para obtener información. Cada vez más se están usando nuevas formas de representación de un conocimiento, distintas del texto lineal impreso. Estos nuevos tipos de publicación digital (desde el texto, hasta una grabación sonora o de vídeo, pasando por un mensaje de correo electrónico, el código fuente de un programa de ordenador para el cálculo de resultados, etc.) han multiplicado las formas de comunicación científica, involucrando un proceso sin precedentes que algunos denominan ya *reformulación del sistema literario*³ y que implica, como dice Piscitelli, *una metamorfosis tecnocognitiva*⁴. La WWW es en la actualidad una de las principales fuentes de información, un vasto depósito informativo que cubre múltiples materias y áreas de aplicación.

El fenómeno de la Web, sin perjuicio de otras interpretaciones, suele asociarse a la infraestructura tecnológica que la hace posible. Si bien es cierto que es necesario un desarrollo y evolución de la infraestructura tecnológica⁵, de los mecanismos

³ Tatiana V. Ershova & Yuri E. Hohlov. Migrar de la biblioteca de hoy a la biblioteca de mañana: ¿Re-o E-volución? En: *IFLA Council and General Conference (66. 2000. Jerusalem)* [documento HTML]. IFLA, 29 de julio de 2000. Disponible en: <http://www.ifla.org/IV/ifla66/papers/063-110s.htm> (Consultado el 1 de septiembre de 2000). Estas autoras aluden al pensamiento de Geoffrey Nunberg, profesor de la Universidad de Stanford, quien afirma que los nuevos formatos, tales como el multimedia y el hipertexto, nos llevan "más allá del libro" creando nuevos modos de lectura y nuevas formas de intercambio intelectual y cultural, que hacen de la comunicación científica no un compendio estático de información, *sino más bien interfaces* [sic.] *dinámicas hacia un discurso de final abierto*.

⁴ Alejandro Piscitelli. *Ciberculturas. En la era de las máquinas inteligentes*. Buenos Aires: Paidós, 1995, p. 19. Este filósofo argentino denomina "metamorfosis tecnocognitiva" a *la causalidad recíproca entre la metamorfosis de los modos de comunicación y la estructuración de la percepción*. El cambio que está sufriendo la comunicación gracias a Internet provoca una nueva aprehensión del conocimiento, lo que no dice Piscitelli es que esa nueva forma de percibir la información, demanda nuevas formas de organizarla.

⁵ La ingeniería de la información necesita mantener una visión clara de las opciones y de las necesidades con las que se puede encontrar en el desarrollo de la Sociedad de la Información, ya que ésta crece a un ritmo vertiginoso, aparecen nuevos mercados y se generan nuevas necesidades tecnológicas. En la actualidad no hay una perspectiva clara sobre el futuro de las tecnologías, pero es obvio que se necesitarán nuevas infraestructuras de información.

principales tales como el protocolo IP (*Internet Protocol*), de forma que se asegure el crecimiento futuro de Internet, también es necesaria una infoestructura coherente que asegure no sólo el acceso, sino también la recuperación y el intercambio de información. Recuperar información relevante resulta cada vez más difícil desde el punto de vista de que el conocimiento humano ya no está centralizado en unidades físicas de información, sino que se ha convertido en un acervo distribuido de informaciones con formatos diversos y heteróclitos. Todas estas formas de información electrónica distribuida a través de la WWW, reciben la denominación espontánea de **biblioteca digital, electrónica** e incluso, **virtual**, en un contexto donde cada vez existen más documentos primarios completos en formatos de codificación numérica.

Los recursos electrónicos son ya una parte integrante de las colecciones de muchas bibliotecas e incluso protagonizan muchos de los servicios. Pero *la Red* (Internet) *no es una biblioteca digital*⁶, no obstante, continuará creciendo y prosperando como nuevo medio de comunicación, por lo que será necesario hacer mucho más de lo que se ha hecho en los servicios bibliotecarios tradicionales para organizar, acceder y preservar la información de la Red. Diseñar un acceso robusto a la información, tanto a la información digital local, como a los recursos electrónicos accesibles en la Web, se ha convertido en un tema de crucial importancia. Al mismo tiempo, las bibliotecas se están convirtiendo en *portales* de información para sus usuarios y planifican arquetipos de "biblioteca virtual" a tenor de la propagación de la información electrónica. El concepto de biblioteca digital encierra, al menos, tres realidades importantes: 1) las bibliotecas que basan la formación de sus colecciones

Los aspectos relacionados con la infraestructura de la Web exceden los objetivos de esta investigación, sin embargo es un tema estrechamente vinculado, como veremos, a la interoperabilidad entre metadatos y al intercambio de la información. Para cuestiones relacionadas con la infraestructura de Internet, *Vid.* Information Infrastructure Futures and Information Society [documento HTML]. *El.pub*, rev. 22 de noviembre de 2000. Disponible en: <http://www.elpub.org/top029.htm> (consultado el 10 de enero de 2001).

⁶ Clifford Lynch. Searching the Internet [documento HTML]. *Scientific American*, March 1997. Disponible en: <http://sciam.com/0397/issue/0397lynch.html> (consultado el 8 de julio de 1998).

en la digitalización de materiales convencionales convirtiéndose en editoras y custodias de su colección, 2) las que se basan en la selección de recursos electrónicos distribuidos organizando su acceso, y 3) las que tratan de integrar objetos de información digital con sus colecciones preexistentes en formato tradicional.

Para poner en práctica el modelo de biblioteca digital como nuevo sistema de comunicación del conocimiento, distintas comunidades de investigación y diversos colectivos relacionados con la información están colaborando, persiguiendo un doble objetivo: por un lado, diseñar, construir y evaluar los mecanismos que permitan integrar en un recurso de información único, las colecciones multiplataforma heterogéneas, diseñadas inicialmente con distintos propósitos y que pertenecen a instituciones diferentes distribuidas en la WWW; y por otro lado, hacer esos recursos digitales accesibles y que permitan que cualquier usuario, de cualquier lugar y en cualquier momento, pueda recuperar información de calidad.

Los metadatos, ante esta realidad, constituyen un mecanismo crítico tanto en la representación del conocimiento de colecciones digitales como en la explotación y aprovechamiento de los datos. Son elementos o estructuras de organización de la información que, asignados a cada objeto de información electrónica, la clasifican, categorizan o describen. Son en definitiva, "datos sobre datos", información estructurada sobre la información distribuida, que se muestran como una de las posibilidades para optimizar la recuperación de información en la nueva Era Digital. La idea de crear identificadores, estructuras y localizadores que organicen, sistematicen y permitan ubicar la heterogénea información albergada en Internet, es lo que ha llevado a que proliferen estándares, esquemas o tentativas de metadatos diferentes, pero con una finalidad semejante: identificar y describir los recursos de información digital, teniendo en cuenta, sus condiciones de uso, la valoración de los documentos electrónicos, su autenticación, su preservación y sobre todo los patrones bibliográficos que fomentarán su recuperación e intercambio. A tenor de las nuevas formas de organización del conocimiento y desechada la quimera de la biblioteca universal única, se ha empezado a apostar y a creer en una Web semántica y en la interoperabilidad entre metadatos. Por ello, a pesar de otras aplicaciones que

puedan tener, este trabajo de investigación se centrará en estudiar y discutir los metadatos como medio para mejorar la recuperación de información en la Red, en concreto, en el ámbito de las nuevas bibliotecas digitales. Para este planteamiento, partimos de la hipótesis de que hoy en día la utilidad de la metainformación sólo puede asegurarse en lo que nosotros denominamos "contextos finitos de información" por contraposición a la inabarcabilidad de toda la Web.

Las bibliotecas necesitan trabajar *con* y no *contra* Internet. Desde que Internet se ha convertido en una fuente reconocida, aceptada e insustituible para la difusión de la información, los bibliotecarios y documentalistas de todo el mundo estamos trabajando para mejorar los métodos de descripción, organización y recuperación de documentos de acceso remoto en línea así como de otros objetos digitales. Es como si muchos bibliotecarios hubiésemos despertado y, de la noche a la mañana, estuviésemos sufriendo una metamorfosis kafkiana en informáticos —que diría Dovey en algunas de las fuentes que citaremos—. Por otra parte, los profesionales de la información no estamos solos en esta empresa. Otros sectores de nuestra sociedad como los creadores y proveedores de recursos electrónicos, las empresas de software, los desarrolladores de servicios de información en línea, etc., están comprometidos en el intento de gestionar y controlar este pródigo mundo de información digital. Esta eventualidad enriquece, a la par que dificulta este trabajo, ya que conlleva el desarrollo simultáneo de una amplia variedad de esquemas y estándares de metadatos. Cada grupo, cada colectivo que maneja y valora la información —por supuesto, el sector de las bibliotecas entre ellos— plantea su propia aproximación al problema de la organización y el acceso a la información electrónica desde su marco particular de referencia e incluye, en los modelos de metadatos que definen, el tipo de información que mejor representa las necesidades de la comunidad de usuarios a la que va dirigida. Por ello, nos interesa determinar también si la proliferación de esquemas de metadatos mejorará la descripción y el acceso a los usuarios, o simplemente comportará nuevos problemas y retos en el ámbito de la normalización y la interoperabilidad, tanto en entre los distintos modelos de metadatos, como entre

éstos y el MARC (*Machine Readable Cataloging*), en el que se funda la mayor parte de los catálogos bibliotecarios preexistentes.

Tradicionalmente, la biblioteca ha representado el deseo de recuperar la información en el *texto fijo múltiple producto de la tecnología de la imprenta*⁷, promoviendo el desarrollo de bases de datos bibliográficas y lenguajes controlados para establecer vínculos jerárquicos y asociaciones entre los términos que se usan para describir el contenido del texto, y para crear sistemas de búsqueda sobre éste. Sin embargo, los nuevos documentos electrónicos se caracterizan principalmente porque permiten la búsqueda en texto completo, de tal forma que el propio documento es al mismo tiempo el continente del texto y la herramienta de acceso intelectual a su contenido⁸. Además se añaden otros objetos de información accesibles en red que conforman un entorno informativo aún más complejo y plural, pero sobre todo digital.

La BIBLIOTECA DIGITAL o VIRTUAL, en mayúsculas, es aún hoy, y será siempre desde nuestro punto de vista, una entelequia⁹, pero las "pequeñas" bibliotecas digitales/virtuales¹⁰ son ya una realidad provocada por Internet. Las

⁷ George P. Landow. *Hipertexto: la convergencia de la teoría crítica contemporánea y la tecnología*. Barcelona: Paidós, 1995, p. 32.

⁸ Hervé Le Crosnier. Les bibliothécaires et le réseau. Un métier qui évolue avec les technologies. En: *Les nouvelles technologies dans les bibliothèques*. Michèle Rouhet, dir. Paris: Electre – Éditions du Cercle de la Librairie, 1996, p. 351.

⁹ Aunque hay quienes —los "*evangelistas electrónicos*" como ironiza Oddy — creen en el hecho cierto y seguro de que Internet es la biblioteca del futuro. [...] *The electronic evangelists believe that all this points to the sure and certain fact that the Internet is the library of the future*. Pat Oddy. *Future Libraries: Future Catalogues*. London: Library Association Publishing, 1996, p. 5. Internet en sí misma no es la biblioteca del futuro, pero evidentemente marcará el futuro de las bibliotecas.

¹⁰ En este sentido, se pueden comentar muchos casos de bibliotecas digitales como por ejemplo: la *Alexandria Digital Library*. <<http://www.alexandria.ucsb.edu>>, uno de los seis proyectos de la DLI (*Digital Library Initiative*) de la *National Science Foundation (NSF)* cuya finalidad es diseñar, implementar y desarrollar una biblioteca digital donde la colección y los servicios se centran en información geoespacial; o la biblioteca que hemos elegido para volcar nuestra tesis en un proyecto de aplicación real: la *Biblioteca Virtual Miguel de Cervantes* <<http://www.cervantesvirtual.com>> que, si bien hoy en día no basa su recuperación en metadatos podría hacerlo en el futuro. *Vid.* Capítulos 8 y 9.

nuevas formas de edición, la edición electrónica y multimedia, han hecho estremecer, en cierto sentido, no sólo al mundo editorial sino también al mundo de los derechos de autor y sobre todo, y desde la perspectiva que se establece en esta tesis, al mundo bibliotecario y a la recuperación de información. A pesar de que reconocemos estos cambios e incluso, en algunos planteamientos de este trabajo, partimos de ellos, también queremos sustraernos de la tendencia, que se advierte en algunos sectores profesionales, a frivolar en exceso con la irrupción de Internet en la escena documental. La información electrónica nos puede llevar a pensar en una redefinición de la misión de la práctica bibliotecaria. No obstante, creemos que la verdadera evolución de las bibliotecas digitales se debe construir sobre el terreno sólido de la práctica acumulada en las bibliotecas tradicionales. Las bibliotecas y el profesional de la información han aceptado todas estas nuevas tendencias tecnológicas como un desafío para ampliar la comunidad de usuarios, pero han tenido y tienen que realizar importantes esfuerzos para incorporar todo este universo de información electrónica a las bibliotecas tradicionales, estudiando, evaluando y categorizando los recursos de Internet en aras a atender las expectativas de los nuevos usuarios de estos incipientes servicios de información.

Esa idea previsor de un nuevo espacio digital entendido como biblioteca (con pautas para una indización que facilite la descripción del contenido, con ayudas para la recuperación y con las referencias bibliográficas de los materiales relacionados en el nuevo entorno) puede defraudarnos. Para evitar la desilusión de ese incipiente espacio electrónico como sistema de información, se están desarrollando diferentes soluciones abocadas a mejorar la recuperación de información en Internet. Una de esas soluciones, en la que basamos esta tesis doctoral como hemos dicho, es el desarrollo de estándares de metadatos. Estos modelos de descripción de la información difieren en sus niveles de especificidad, estructura y madurez, así como en su propia terminología, sin embargo su objetivo principal es el mismo: crear estructuras de base que sirvan para describir, identificar y definir los distintos objetos de información distribuidos en la Web, de tal forma que la búsqueda basada en ellos

solucione, en cierta medida, la recuperación, valoración, preservación y acceso global a los recursos de la Red.

El problema de la recuperación de información —perspectiva desde la que abordamos el tema de los metadatos en esta investigación— no es reciente, ni está exclusivamente vinculado al entorno de la información electrónica o de Internet; podríamos decir, sin que fuese desmedido, que es tan antiguo como las bibliotecas. Desde los catálogos manuales hasta las bases de datos bibliográficas, la recuperación de información ha sido siempre objeto de investigación, pero ha recobrado un nuevo interés con el desarrollo de la comunicación en red y con la multiplicación de contenidos en formato electrónico distribuido en el espacio Web, reverberando por ello las técnicas de búsqueda y recuperación tradicionales y haciendo surgir otras nuevas.

La recuperación de información automatizada se ha estudiado tradicionalmente atendiendo a tres aspectos principales: *estructuras de información, análisis de contenido, y evaluación*¹¹, observados a su vez, desde una doble perspectiva. Por un lado un enfoque tecnológico o algorítmico y por otro, una perspectiva semántica o cognitiva. Ante el problema de la recuperación de información en Internet, podemos identificar igualmente dos tendencias similares:

- Una tecnológica, protagonizada por el perfeccionamiento de las técnicas de indización automática en texto completo, el procesamiento del lenguaje natural y el desarrollo de nuevos algoritmos de búsqueda y recuperación de información. En esta tendencia podemos destacar la insuficiencia de los motores y servicios de recuperación de información actuales, debida fundamentalmente a la heterogeneidad de los datos distribuidos en la Red, a

¹¹ Rijsbergen y otros autores clásicos de la teoría de la recuperación de información (Salton, Cleverdon, etc.) han señalado estas tres áreas principales de investigación en este ámbito. C. J. van Rijsbergen. *Information Retrieval*. 2nd ed. London: Butterworths, 1979 [documento HTML]. Glasgow: University of Glasgow, Computing Science, rev. 18 de septiembre de 1996. Disponible en: <http://www.dcs.glasgow.ac.uk/Keith/Chapter.1/Ch.1.html> (consultado el 27 de julio de 1998).

la propia naturaleza fragmentada de la indización, a la cobertura incompleta y a su propia incompatibilidad.

- Una tendencia semántica, a la que pertenecería el desarrollo de modelos de metadatos para estructurar la información y, la implementación de vocabularios controlados y listas de autoridad para organizar el contenido de esos metadatos. Bajo este enfoque semántico están surgiendo nuevos sistemas de recuperación de información vinculados a áreas del conocimiento concretas, denominadas genéricamente *subject gateways*, que basan la formación de sus "colecciones virtuales" en la selección de recursos, y la búsqueda, en esquemas de metainformación estructurada.

Por analogía a esta doble perspectiva en el estudio de la recuperación de información automatizada, en esta investigación se muestran también dos tendencias o entornos informativos para los metadatos:

- a) Por un lado, el de los servicios de indización y búsqueda en Internet que basan su funcionamiento, en la mayoría de los casos, en la potencia del software. En este sentido, sólo podemos determinar la funcionalidad de las *metaetiquetas* HTML y del marcado de procedimiento de este lenguaje para aumentar la visibilidad y la promoción de los recursos y páginas Web en sistemas genéricos de recuperación de información en Internet.
- b) Por otro lado, los contextos abarcables o finitos de información, sea su finitud debida al límite físico (por ejemplo una Intranet), al límite de gestión local de acceso global (como en el caso de una biblioteca digital propiamente dicha), o al límite temático o basado en la selección (como en las *subject gateways* o las *clearinghouses*¹²). En estas bibliotecas digitales o sistemas de información es donde realmente existe un control sobre las colecciones y donde se puede basar la recuperación en sistemas de metadatos.

¹² Estos conceptos en inglés se mencionarán y explicarán en distintos apartados a lo largo del trabajo. Para el esclarecimiento de su significado en el ámbito de esta investigación. *Vid.* Glosario.

Estos planteamientos nos permiten abordar la recuperación de información en la Web desde una perspectiva semántica y de cariz bibliotecario, vinculada a la necesidad de establecer metainformación estructurada como alternativa en la organización de colecciones de objetos de información digital. Reconocemos no obstante, que la solución al problema de la recuperación, debe exigir el desarrollo de la investigación en ambos sentidos, tanto semántico como tecnológico, contemplando por ello, el desarrollo de estructuras semánticas de metainformación y la potencia de procesamiento informático de las mismas. Es más, creemos que el desarrollo de las nuevas aplicaciones informáticas debe de contemplar la aptitud casi hermenéutica de las mismas, en el sentido del arte de interpretar recursos Web a los que se ha conferido una semántica gracias a la capacidad y flexibilidad de lenguajes como XML (*eXtensible Markup Language*) y RDF (*Resource Description Framework*).

Es importante señalar además que, a pesar de que el tema de los metadatos haya resurgido en el contexto de Internet, el trabajo bibliotecario y documental tradicional, siempre se ha preocupado por normalizar "metadatos". Es decir, establecer estándares para los datos descriptivos del documento primario a fin de mejorar su recuperación ulterior. Desde las propias reglas de catalogación, el desarrollo de formatos bibliográficos de intercambio como ISO 2709, hasta los distintos vocabularios de control terminológico, siempre han existido distintos datos sobre datos que pretenden describir e incluso sustituir la fuente original impresa de tal forma que anticipen su contenido y/o interés para el usuario antes de su recuperación. Las bibliotecas usan el formato MARC desde los años 60 como estándar para compartir información y para identificar, describir y proporcionar un acceso coherente a sus colecciones. De igual forma, los profesionales que desempeñan su trabajo en archivos y museos han desarrollado a lo largo de su labor profesional sus propias normas descriptivas para preservar, organizar, y proporcionar acceso al patrimonio artístico y documental. Sin embargo, en el contexto de las nuevas e incipientes bibliotecas digitales, y a tenor de la globalización de la información, se están reduciendo las distintas demarcaciones documentales (archivos, bibliotecas, museos, etc.) y es preciso acceder a colecciones diversas y

heterogéneas que necesitan sistemas normalizados entendibles no sólo por los investigadores y usuarios, sino también por la máquina. Dotar a los registros electrónicos de semántica interpretable por los ordenadores será una de las funciones de los metadatos aplicados al mundo digital.

La falta de una normalización sólida y universal, así como la versatilidad de la indización de la información electrónica son obstáculos inminentes para un acceso fácil y productivo a estas futuras-actuales bibliotecas digitales. Esta tesis representa una visión genérica de toda esta problemática, tratando de revelar el *statu quo* de los metadatos como una alternativa para solucionar el problema de la recuperación de información en la Red y como modelo socio-técnico para la organización y el procesamiento de la información en las bibliotecas digitales, que puede protagonizar el cambio de paradigma en la gestión de la información electrónica. Se abrirán aquí por ello, numerosas expectativas y líneas de investigación nuevas.

CAPÍTULO 2:

OBJETIVO, MÉTODO Y FUENTES UTILIZADAS

Una vez planteado en líneas generales el ámbito de esta investigación, este segundo capítulo tiene un triple propósito: por un lado, definir el marco y objetivos de la tesis, fijar los límites de la investigación y justificar la hipótesis de partida. Por otra parte, este capítulo, fundamental en un trabajo de estas características, tiene la finalidad de señalar los diversos aspectos metodológicos (descripción del desarrollo de la línea argumental del trabajo, aspectos relacionados con el aparato crítico del mismo o con la experimentación llevada a cabo) y otras observaciones que conlleva la naturaleza del tema que nos ocupa. Por último, trataremos de analizar las fuentes y la bibliografía utilizadas para describir sus particularidades y medir su importancia a la hora de avalar nuestros planeamientos.

2.1. Marco y objetivos del trabajo

La mayor parte de las investigaciones, propuestas, soluciones, etc. que se desarrollan actualmente en el campo de la Biblioteconomía y Documentación, están estrechamente ligadas a los cambios tecnológicos e incluso, en muchas ocasiones, supeditadas a ellos. Los profesionales de la información debemos enfrentar los retos que supone la tecnología y el compromiso que conlleva para nosotros la Sociedad de la Información. Por todo ello hemos elegido un tema de actualidad (incluso "de moda", como nos referiremos a lo largo de este trabajo): los metadatos y su implicación y aplicación, sobre todo en el nuevo contexto de las bibliotecas digitales.

Se puede decir que los metadatos y las bibliotecas digitales, conceptos principales de esta investigación, están todavía en fase de construcción técnica, e inclusive muchos de los aspectos teóricos están aún sin resolver o, al menos, poco asentados en la comunidad científica. Son muchos los autores que reconocen este hecho¹. Los estándares de metadatos destinados a la World-Wide Web están todavía en desarrollo y continuarán, sin duda, su proceso de evolución. Por su parte, las bibliotecas digitales presentan una casuística poco homogénea y llena de contradicciones, lo que revela también su estado incipiente. Toda esta realidad, unida a la necesidad de instituir una línea de investigación en español sobre este tema, son los condicionantes principales que nos han llevado a enfrentar este trabajo.

El estado embrionario de la materia que tratamos ha dificultado esta investigación, a la par que ha favorecido el resultado de la misma, al desarrollarse en un periodo de efervescencia y discusión. Ha sido complicado, sin embargo, recoger todos los aspectos que confluyen en un tema tan amplio y complejo, teniendo la sensación de tener que asumir el reto de consignar un "macro-estado" de la cuestión, que conciliase, por un lado toda la práctica bibliotecaria tradicional, y por otro el vertiginoso cambio tecnológico y los desafíos que todo ello conlleva. En la bibliografía específica, encontramos información poco arraigada, confusa en algunos casos y complementaria en otros, pero siempre en continua y ágil revisión, que ha mantenido en nosotros un compromiso de seguimiento constante de los recursos de Internet sobre estas cuestiones. Al mismo tiempo, esto nos ha permitido contribuir con aportaciones al estado actual de los metadatos y las bibliotecas digitales en el momento en que se estaban gestando como teoría y como principio de realidad

¹ V. gr. Brisson dice: *The tandem concepts of digital libraries and metadata are currently in their formative stages and for this reason their application still undeveloped [...]* Roger Brisson. *The World Discovers Cataloging: A Conceptual Introduction to Digital Libraries, Metadata and the Implications for Library Administrations. Journal of Internet Cataloging*, 1999, vol. 1, nº 4, p. 5. Alegatos como éste ratifican la oportunidad y la necesidad de investigación en estos temas, y por tanto, los objetivos de esta tesis doctoral.

práctica, pudiendo dar con ello un enfoque mixto a nuestra investigación: analítico-descriptivo y práctico.

Partiendo de este marco, antes de consignar nuestras propias hipótesis, nos parece oportuno evocar la hipótesis hegeliana reelaborada por Blair² que imbuye toda nuestra investigación:

Un crecimiento cuantitativo de cualquier sistema (en este caso, los sistemas de información) provoca cambios cualitativos que tendrán lugar en su estructura y procesos. En este sentido la tarea de la investigación dedicada a la recuperación de información debe de abocarse a descubrir cómo estos grandes sistemas de información actuales difieren de los pequeños sistemas que conocemos mejor, y cómo podemos hacer que estos grandes sistemas sean lo más eficaces posibles.

A pesar de lo genérico de este aserto de Blair, podemos reconocer la vigencia de la hipótesis en cualquier aspecto de la investigación en recuperación de información. La finalidad de traerlo aquí ahora, no es otra que la de trasladarlo al contexto que reflejamos y tratamos en esta tesis doctoral, y podría ser la hipótesis extensiva y palmaria de cualquier investigación en Documentación que, como ésta, asuma el reto de la información electrónica de Internet. La cantidad de información accesible en el nuevo e inabarcable "sistema de información" de la World Wide Web conlleva cambios cualitativos en lo relativo a la organización y acceso a la información. Por ello, las nuevas investigaciones en recuperación de información deben dedicarse a dirigir estos nuevos sistemas cimentados en la amplitud de Internet, a una gestión eficaz y eficiente de la información, partiendo, en ciertos aspectos, de los principios conocidos y dominados en el trabajo bibliotecario secular, sin que esto impida plantear nuevas propuestas. Así, se barajarán por un lado conceptos "tradicionales" como biblioteca, normalización, proceso técnico o control documental, que han preocupado a los bibliotecarios y documentalistas durante

² David C. Blair. *Language and Representation in Information Retrieval*. Amsterdam, etc.: Elsevier Science Publishers, 1990, p. IX.

siglos, y por otra parte se hablará de lenguajes de marcado, bibliotecas digitales, motores de búsqueda, interoperabilidad, o el propio concepto de metadatos, que han surgido a tenor de la explosión de la Web y de la información accesible en red. La imbricación de todos estos problemas nos ayudará a demostrar que la recuperación de información eficiente en Internet debe ser fruto de un continuo perfeccionamiento tanto de las técnicas de organización de información, como de la tecnología.

Sostenemos pues, que la gestión eficaz de la información en la WWW dependerá de una gestión normalizada de estructuras de organización de la misma (metadatos). Según esto, trataremos de acercarnos a la idea de que la caracterización o descripción de los documentos electrónicos por medio de metadatos se plantea como una solución para mejorar la relevancia en la recuperación de información en Internet, no sólo textual sino también de imágenes o de cualquier otro tipo de objeto de información electrónica. Como adelantábamos en la introducción de este trabajo, son muchas las preguntas que nos han llevado a abordarlo, *v. gr.*: ¿Los metadatos son una moda o una necesidad? ¿El trabajo bibliotecario tradicional está basado en metainformación? ¿Son los metadatos la respuesta a la nueva gestión de la información en Internet? ¿Cómo, quién y cuándo se deben crear metadatos? ¿Pueden estas estructuras de datos sobre los datos mejorar la recuperación en toda la Red o simplemente son el bastión de sistemas de información finitos organizados *ad hoc*? ¿Qué iniciativas se están desarrollando y qué validez e integración tiene cada una de ellas? ¿Qué papel juega la normalización en todo este contexto? ¿Son las bibliotecas digitales el marco de aplicabilidad real de la metainformación? ¿Qué nuevos retos implica la gestión de la información global?, etc. A éstas y a otras muchas cuestiones trataremos de dar respuesta a lo largo de esta tesis doctoral, integrando todas ellas, al final, en un proyecto de aplicación real.

Teniendo en cuenta estas hipótesis, los **objetivos** concretos que persigue esta investigación, y que anticipan algunas de las reflexiones llevadas a cabo en ella, son:

1. Realizar un análisis minucioso y exhaustivo de la bibliografía científica relativa al concepto y desarrollo de metadatos, perfilando los distintos entornos en los que

se usan, haciendo hincapié en la evidencia de que en un espacio de datos infinitamente amplio, la gestión eficaz de la información en red —y por ende su recuperación— dependerá, cada vez más, de la gestión eficaz de metadatos.

2. Reconocer y señalar las diferencias fundamentales entre la actividad catalográfica tradicional y los metadatos, así como la relación que existe entre los metadatos y los objetos a los que se refieren, y los registros catalográficos y los objetos que referencian.
3. Estudiar y determinar cuál es la forma más idónea de asignación de metainformación, quién debe ser el agente que cree los metadatos, así como hacer una revisión de las herramientas y aplicaciones de software desarrolladas a tal efecto, distinguiendo aquellas destinadas a mejorar la visibilidad en los motores de búsqueda, de aquellas desarrolladas para crear metadatos, en sentido estricto, para la recuperación de información en bibliotecas digitales.
4. Analizar y evaluar las propuestas e iniciativas de modelos de metadatos más significativas en el contexto de esta investigación, distinguiendo los esquemas de propósito general, que nacen con una vocación normativa para toda la Web, de los formatos con propósitos específicos, que no son otra cosa que reductos de aplicabilidad del concepto y forma de los metadatos a un dominio informativo particular.
5. Considerar el problema de la estandarización en el ámbito de los metadatos y de las bibliotecas digitales en una Sociedad de la Información Global, analizando las ventajas e inconvenientes de la normalización, así como los procesos de formalización de normas para la descripción y recuperación de recursos electrónicos.
6. Considerar el proceso de desarrollo y las tendencias de los nuevos sistemas de información basados en el documento digital y en la edición electrónica, donde los documentos se entienden como objetos, así como los problemas que plantea la búsqueda y recuperación de información en ellos.

7. Distinguir el uso de metadatos en la recuperación genérica en Internet (a través de directorios, motores de búsqueda, metabuscadores, etc.), de su utilización, más factible y controlable, para organizar y recuperar información en un conjunto abarcable de recursos electrónicos que constituyen bibliotecas digitales y/o sistemas de información temática de calidad.
8. Destacar los problemas tradicionales de la recuperación de información documental, magnificados por el universo de información heterogéneo, multidisciplinar y multilingüe de la Red, no sólo en cuanto a la estructuración de la información sino también en lo relativo al control del vocabulario y la necesidad de desarrollar estándares y estructuras que soporten el uso compartido de herramientas de organización del conocimiento en entornos de red.
9. Señalar los riesgos y oportunidades que supone la utilización de estándares de metadatos para de descripción de objetos documentales, tanto desde el punto de vista técnico, como de las implicaciones que subyacen para el desarrollo de colecciones electrónicas, y de la redefinición del papel del profesional de la información en este contexto.
10. Proponer una definición de biblioteca digital a tenor de la organización, a través de sistemas de metadatos, de los documentos entendidos como objetos de información (DLOs). Distinguiendo las bibliotecas basadas en la digitalización de documentos tradicionales que conforman una colección digital real, de las constituidas por colecciones de recursos originales, y precisando el proceso de creación de metadatos.
11. Analizar el nivel de normalización y/o interoperabilidad, real y exigible, para esta forma de estructuración y organización de la información (metadatos), tanto desde el punto de vista de la homogeneidad del esquema, como del control del contenido, de tal forma que se pueda retomar el planteamiento de biblioteca digital global y de bibliotecas híbridas locales.
12. Plantear el diseño de un sistema de información basado en metadatos para el caso concreto del mayor proyecto español de biblioteca digital: la Biblioteca Virtual

Miguel de Cervantes (BVC). Este aspecto práctico de la tesis tiene, a su vez, sus propios objetivos:

- Estudiar la composición, las características de la colección, la tipología de documentos, las necesidades, los objetivos, etc. de la BVC en aras a hacer un diagnóstico para la aplicación de metadatos.
 - Destacar las particularidades de la BVC como productora y gestora de información digital, y sus flujos de trabajo en ambos sentidos.
 - Proponer un modelo de gestión de metadatos teniendo en cuenta las características de la información que alberga, así como la estabilidad de los estándares de metadatos existentes hasta el momento.
 - Plantear la viabilidad de un proyecto de aplicación de metadatos, pionero en el ámbito español de las bibliotecas digitales, donde se puedan poner a prueba todas las reflexiones que implica esta tesis y mostrar la eficacia de esta tecnología para la recuperación de información en un contexto determinado de biblioteca digital.
13. Reflexionar en general, sobre la terminología y la designación de nuevos conceptos en torno a los sistemas de información en red y los metadatos, tratando de resolver el vacío designativo en español de algunos conceptos relacionados con el tema de este trabajo.
14. Por último, instituir un punto de partida teórico-práctico para la investigación en metadatos y bibliotecas digitales, sentando las bases de una investigación aplicada posterior a esta tesis doctoral, tanto en la Biblioteca Virtual *Miguel de Cervantes* como en otros proyectos.

En cuanto a los límites de este trabajo, es preciso aclarar que se han excluido deliberadamente algunos temas relacionados que, desde nuestro punto de vista, requieren un tratamiento específico y darían lugar a investigaciones monográficas de características semejantes a ésta. No obstante, a lo largo de la tesis haremos mención expresa a ellos para avalar algunas de nuestras hipótesis. Estos temas son, fundamentalmente dos:

a) El nivel de uso de metadatos. Aunque haremos alusiones frecuentes a la implantación de la metainformación en la WWW, no incluimos un análisis genérico y sistemático sobre el uso de etiquetas <META>³. Las razones de no introducir en este trabajo un estudio concreto de su utilización son las siguientes:

- Los muestreos que hemos hecho en este sentido siempre corroboran la escasa utilización de metaetiquetas y/o metadatos en páginas Web.
- A pesar de que existen herramientas de software que permiten analizar la inclusión de metaetiquetas con mucha facilidad⁴, sólo son aplicables cuando los metadatos están embebidos en el código HTML del documento. Por ello, si tratásemos de analizar exhaustivamente su utilización, los resultados siempre serían parciales y reflejarían exclusivamente el uso de metadatos en páginas y sitios Web, donde el único sistema de recuperación al que están sometidos son los buscadores genéricos de la Web. No quedaría reflejada su implantación en bibliotecas digitales concretas, donde los metadatos pueden y suelen estar recogidos en otros sistemas de almacenamiento de información como bases de datos, que no tienen por qué tener necesariamente un reflejo en el código fuente HTML.

³ En general a lo largo de todo el capítulo 4, concretamente en la nota 1, se citarán trabajos que demuestran el nivel de utilización de metaetiquetas en las páginas y sitios Web de las bibliotecas españolas. Asimismo resaltaremos las funcionalidades de las herramientas de software en este sentido, de tal forma que quede patente que, a pesar de que no realizamos ningún muestreo concreto de la utilización de metadatos, sí apuntamos la metodología y las herramientas informáticas para realizarlo. También serán frecuentes las alusiones a la utilización de metaetiquetas al comentar la ausencia de masa crítica en su aplicación en los primeros años del DC y al reflexionar sobre la evolución en la utilización de metadatos sencillos o etiquetas <META> del HTML por los motores de búsqueda genéricos de la WWW (Vid. 7.2.).

⁴ Algunas de ellas las analizaremos también en el capítulo 4, p. ej. Metabrowser facilita la visualización de los metadatos de forma simultánea a la navegación (Fig. 41), MetaTag Maximizer realiza un diagnóstico sobre la utilización de metaetiquetas (Fig. 22) y Metabot permite la generación de informes sobre el uso de metadatos, a partir de una selección de elementos en virtud de los cuales se quiere realizar el análisis (Fig. 44). Vid. 4.3.

- Además, la tendencia latente en toda la tesis es proponer un modelo de utilización de metadatos en un entorno concreto de información electrónica, concebido como biblioteca digital, no un modelo de recuperación general sobre toda la Web. Esto justificaría un análisis de metainformación de los recursos que forman esas bibliotecas digitales que, en el caso de no estar embebidos en el documento, no permitirían el acceso a su análisis, o que en muchos casos las etiquetas <META> del código fuente de visualización en el navegador (HTML) no responderían a la estructura profunda de metadatos (en RDF, XML, etc.) que fundamenta la recuperación de información en ellas.
- b) Otro de los temas excluidos de este trabajo es el problema de determinar los recursos a los que se deben asignar metadatos. El hecho de no afrontar con profundidad este aspecto tiene su justificación en que se trata de un tema de complejidad y entidad individual suficiente para protagonizar una investigación monográfica. Abordar este aspecto implicaría realizar un estudio sobre la evaluación de recursos Web y el control de calidad de las colecciones formadas por documentos electrónicos, lo que excede claramente, los objetivos específicos de esta investigación.

Finalmente, aunque no lo excluimos *ex professo*, no tratamos en detalle la metainformación relacionada con el contenido temático de los documentos (*subject based meta data*) esto es, los metadatos que se refieren a datos que representan materias y sus interrelaciones, protagonizados por esquemas como SHOE (*Simple HTML Ontology Extensions*), XOL (*Ontology Exchange Language*), XTM (*XML Topic Maps*), o esquemas basados en RDF como OIL (*Ontology Interchange Language*). Sin embargo sí trataremos la importancia de este tipo de metadatos y nos referiremos a su normalización en el apartado 6.4., partiendo del punto de vista más tradicional en Documentación, los tesauros, pero teniendo en cuenta su proyección y visualización en la Web como estándares de contenido de metadatos.

2.2. Aspectos metodológicos

La metodología de elaboración de este trabajo de investigación es eminentemente inductiva en lo que se refiere al desarrollo del tema, pues a lo largo del mismo se estudian modelos, proyectos, herramientas, casos de aplicabilidad, estándares, problemas, etc. relacionados con los metadatos, para llegar a la formulación general de la hipótesis de que, en la actualidad, la recuperación de información a través de metadatos es aplicable en entornos finitos o abarcables de información. Trataremos de demostrar esta hipótesis a lo largo de toda la investigación, especialmente en una última parte, donde volcaremos todas las reflexiones que llevamos en torno a este tema, en un contexto concreto de información Web o biblioteca digital: la Biblioteca Virtual *Miguel de Cervantes*, que puede optimizar su sistema informativo a tenor de la aplicación de estructuras de metainformación entendibles por máquina.

En cuanto al proceso de elaboración en sí mismo, cada parte y/o fase de la investigación tiene su propia metodología. En un principio, y de forma general a lo largo de todo el trabajo, combinamos el reconocimiento y lectura de las fuentes más importantes sobre recuperación de información y bibliotecas digitales relacionadas con los metadatos, con la observación y experimentación de esta tecnología en la propia Red. Para establecer esta metodología nos basamos en la identificación y seguimiento de los grupos más activos o predominantes en esta área —W3C (*WWW Consortium*), UKOLN (*United Kingdom Office for Library and Information Networking* de la Universidad de Bath), DCMI (*Dublin Core Metadata Initiative*), OCLC (*Online Computer Library Center*), el DSTC (*Distributed Systems Technology Centre*) etc.— ya que su actividad genera publicaciones de sus investigadores (como por ejemplo las de los miembros de la UKOLN en la revista *Journal of Documentation*), seminarios de discusión y grupos de trabajo (como los *workshop* de la DCMI), y también proyectos, prototipos y aplicaciones en la Red (como el proyecto *Harmony*, la herramienta de creación de metadatos *Reggie* o el prototipo de recuperación de información *HotMetal*, del DSTC).

Al tratarse de una investigación dilatada en el tiempo ha sido fundamental la continuidad y la actualización de las fuentes y los problemas relativos a los metadatos, requiriendo un seguimiento minucioso de los diversos temas que planteamos a lo largo de la misma⁵. Para llevar a cabo esa renovación constante de los problemas, hipótesis y soluciones que finalmente recogemos aquí, ha sido insoslayable también la consulta y/o suscripción a las diferentes listas de discusión, foros de debate y grupos de *news*. Las listas de distribución y los *newsgroups* constituyen, desde nuestro punto de vista, la exacerbación moderna de los colegios invisibles y, por ello, el debate científico más importante, o al menos más diligente y actualizado, sobre el tema ya que, como se ha mencionado, se trata de una materia en desarrollo, sometida todavía al debate científico, tanto entre informáticos como entre

⁵ Este seguimiento se demuestra, por ejemplo, en el hecho de que en 1998 ya habláramos de XML (*eXtensible Markup Language*) como *la promesa para constituir un verdadero estándar de intercambio de documentos en la WWW*. Eva M^a Méndez Rodríguez. *Metadatos y recuperación de información*. Tesina de doctorado, dirigida por Virginia Ortiz-Repiso Jiménez, presentada en septiembre de 1998, p. 120, (ejemplar original de autor). Por otra parte, en 1999 aludíamos a *las oportunidades que plantea el esquema RDF como "piedra Rosetta" en el mundo de la metainformación legible por ordenador*, ante la necesidad de interoperabilidad entre formatos; y que dicho formato (*Resource Description Framework*) podía *convertirse en el componente esencial de cualquier biblioteca digital efectiva*. Eva M^a Méndez Rodríguez. *RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio*. En: *Jornades Catalanes de Documentació* (7. 1999. Barcelona). Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 1999, p. 495.

Pasado el tiempo, en el momento de presentación de esta tesis doctoral (diciembre de 2001) afirmaciones como éstas, que demuestran una continuidad en la investigación de los temas que tratamos, se han verificado en la práctica. Así, XML es una verdadera norma de intercambio de información y RDF es un formato esencial en los planteamientos de biblioteca digital más serios e innovadores, e incluso se ha convertido en un estándar *de facto* para el diseño de software comercial para la gestión de bibliotecas digitales. Por ejemplo RDF es el formato de representación e intercambio de información en herramientas de gestión de bibliotecas digitales, como las de la empresa Ex-Libris (*Cfr.* 4.3.2.3. *DigitoolLibrary*, capítulo 4, nota 213), o en proyectos europeos para la integración de bibliotecas digitales como Renardus. (*Vid.* Lesly Huxley. *Renardus: Fostering Collaboration Between Academic Subject Gateways in Europe*. *Online Information Review*, 2001, vol. 25, nº 2, p. 121-127).

bibliotecarios y profesionales comprometidos en la difusión de la información electrónica⁶.

Esta es una investigación "con" y "para" Internet o, más concretamente, con y para la World Wide Web. Esta asunción ha impregnado todo el trabajo, también desde el punto de vista metodológico y nos lleva a citar e interpretar algunas ideas de su inventor. Tim Berners Lee, hablando de los análisis que se hacen de la WWW, dice que la investigación en línea es un aspecto más serio, *uno trata de encontrar no sólo la respuesta a una pregunta, sino también qué estructuras puede haber en la información*⁷ y eso es justamente lo que tratamos de hacer en este trabajo. Por otra

⁶ Son muchas las listas de distribución generadas en torno al tema de los metadatos. Sólo en el servicio de listas JISCMAIL <<http://www.jiscmail.ac.uk/category/P14.html>> se consignan (12/08/01) veintiséis listas relacionadas exclusivamente con el modelo Dublin Core. No obstante, las principales listas, cuya consulta o suscripción nos ha servido para introducir algunos de los aspectos más actuales relacionados con los temas de esta investigación, así como para contar con un contrapunto interactivo a nuestras propias hipótesis, son las siguientes:

- La suscripción, desde agosto de 1999 a la lista del consorcio Web sobre el *Resource Description Framework*: RDF-interest, aunque en algunos aspectos muy técnica, ha sido fundamental para el seguimiento de la evolución de RDF. Los mensajes públicos se pueden consultar en: <http://lists.w3.org/Archives/Public/www-rdf-interest>; Brian McBride, investigador de HP en Bristol (Inglaterra), ha hecho una selección muy interesante de los temas planteados en torno a RDF en esta lista de discusión con enlaces a los mensajes más interesantes al respecto, *Vid.* <http://www-uk.hpl.hp.com/people/bwm/rdf/issues.htm>
- Un clásico en el mundo de los debates electrónicos sobre metadatos, en especial sobre el modelo Dublin Core, es DC-general. Esta lista inicialmente se llamó META2 y dependía del proyecto ROADS <<http://www.roads.lut.ac.uk/lists/meta2>>. Posteriormente cambió de nombre y pasó al servicio de listas de Mailbase y en la actualidad está en JISCMAIL. El registro de los mensajes emitidos en esta lista se pueden consultar en: <http://www.mailbase.ac.uk/lists/dc-general> hasta el 24 de noviembre del 2000, y para mensajes posteriores a esta fecha, en: <http://www.jiscmail.ac.uk/lists/dc-general.html>
- Por otra parte, la lista NKOS (*Networked Knowledge Organization Systems/Services*), donde el tema de discusión es el establecimiento de modelos de referencia para su uso e interoperabilidad en Sistemas de Organización del Conocimiento en la Red, a la que nos suscribimos desde su creación (principios de julio de 1998), ha enriquecido mucho nuestros planteamientos y ha abierto expectativas de investigación futura en relación con los estándares de metadatos basados en materias y la visión, desde el punto de vista de la documentación (tesauros, ontologías, etc.) de los mismos. Los mensajes desde el 18/09/98, se pueden consultar en: <http://orc.dev.oclc.org:5103/nkos/maillist.html>

⁷ Tim Berners Lee. *Tejiendo la red: el inventor del World Wide Web nos descubre su origen*. Madrid: Siglo veintiuno, 2000, p. 164.

parte, *el Web rompe los límites en los que confiábamos..., pero también puede crear límites nuevos*⁸. Aunque Berners Lee utiliza la expresión para referirse a los límites espaciales que supera la WWW, nos parece interesante reflexionar sobre el sentido abstracto de la idea. Así, podemos argumentar que en esta investigación la Web, como fuente y objeto de estudio a la vez, pone en entredicho los límites tradicionales en ambos sentidos. Por un lado, como fuente de información, convierte en inabarcable la bibliografía, y por otro, como objeto de estudio relacionado con las bibliotecas, crea límites nuevos, como son, verbigracia, el acceso y la recuperación de información, a la par que rompe las fronteras del control documental tal y como lo concebíamos hasta ahora.

Además, dice también Berners Lee, *es importante que el Web ayude a la gente a ser intuitiva además de analítica, porque nuestra sociedad necesita de ambas funciones*⁹, por ello esta investigación se deja imbuir de ambas tendencias analítica, de la realidad de los metadatos, e intuitiva, en cuanto a su proyección de futuro.

Un aspecto metodológico fundamental que entraña un trabajo de esta naturaleza es la necesidad de análisis del software para evaluar su rendimiento y funcionalidad. Este componente ha sido esencial, sobre todo, en los capítulos 4, 7 y 8. En el capítulo 4 (apartado 4.3.), se abordó una metodología de evaluación cualitativa en la utilización de distintas herramientas de asignación de metadatos e indización de documentos electrónicos. La elaboración de esta parte de la tesis implicó la búsqueda, selección, instalación, prueba y clasificación de cada una de las aplicaciones que se analizan¹⁰. Se utilizaron todas, haciendo varias pruebas de

⁸ *Ibid.*, p. 186. (Sobre el género del término *Web* utilizado, *Cfr.* Glosario, *Web.*)

⁹ *Ibid.*

¹⁰ Se consigna siempre una descripción sobre la versión evaluada ya que se trata, en la mayoría de los casos, de software en desarrollo y sus características pueden haber variado en el momento de presentación de este trabajo.

asignación de metadatos adaptadas a sus propias prestaciones de funcionamiento¹¹. Para homogeneizar el estudio y evaluar el rendimiento, con casi todas las aplicaciones se utilizó el ejemplo de creación de metadatos ante el supuesto de esta tesis doctoral concebida como DLO (*Document Like Object*) para la Red. A pesar de lo tedioso de esta metodología y de la longitud que tomará el capítulo 4 por ello, ha sido ineludible ejecutar un estudio práctico del software, ya que el objeto último de esta investigación es la aplicación de un modelo de metadatos a una biblioteca digital, y esto hace necesario el conocimiento de las herramientas que pueden coadyuvar a implementar esta tecnología para la recuperación de información.

Por otra parte, el reconocimiento de la necesidad de metadatos, unido a su complejidad y al desconocimiento acerca de cómo implantarlos y qué herramientas utilizar para un sistema de información basado en ellos, ha llevado a que algunas instituciones —como la *State Services Commission* del Gobierno de Nueva Zelanda¹²—, hagan públicas peticiones de información (*Request For Information*,

¹¹ Salvo en el caso de MetaStar, que fue evaluado a partir del acceso a la información de la base de conocimiento Extranet que tiene la empresa como soporte a sus clientes, a través de un *login* y una *password* cedidas por Blue Angel Technologies. En el caso de otros sistemas (SIG-DLOs, Vid. 4.3.2.3.) que, por no tener acceso a la aplicación en sí misma, sólo mencionamos algunas de las prestaciones difundidas en la información pública (Web) sobre ellos.

¹² Uno de los responsables de la Comisión de Servicios del Estado de este país envió un RFI a la lista de correo sobre RDF del W3C. Kaylene Murdoch. Request for Information (RFI) for a Metadata Management Facility and Search Tool [correo electrónico en lista de distribución]. En: *www-rdf-interest*, 5 de agosto de 2001; 23:39. Disponible en: <http://lists.w3.org/Archives/Public/www-rdf-interest/2001Aug/0017.html> (consultado el 6 de agosto de 2001). Para más información acerca de esta petición, Vid. *Request For Information (RFI): Metadata Management Facility & Search Tool* [documento HTML]. New Zeland e-government, State Services Commission, 2001. Disponible en: <http://e-government.govt.nz/vendor/rfi-list.html> (consultado el 6 de agosto de 2001).

Tanto en el mensaje enviado a la lista de discusión del W3C, como en el RFI publicado en la Web, la Comisión de Servicios del Estado del Gobierno de Nueva Zelanda hace pública la convocatoria solicitando información concerniente a la gestión de metadatos, que incluya todas las funciones relacionadas con la metainformación que serían necesarias para implementar en el portal *E-government Citizens* que se presentará a finales de junio del 2002. Esta comisión esperaba obtener (antes del 24 de agosto del 2001) la información suficiente que les permitiera evaluar la disponibilidad de las soluciones potenciales, el estado actual del mercado, y proporcionar el coste indicativo para presupuestar y planificar los objetivos. Lo que nosotros hemos hecho en el apartado 4.3., con la metodología expuesta arriba, es anticiparnos a una RFI como esta y plantear el *statu quo* del mercado

[cont.]

RFI) o soliciten propuestas de software para implementar metadatos en sus sistemas de información.

En los capítulos 7 y 8, las múltiples pruebas que realizamos fueron sobre el funcionamiento de motores y aplicaciones de búsqueda, fundamentalmente en línea, que contemplan la presencia de metadatos o etiquetas <META>, así como la evaluación de proyectos de bibliotecas digitales en la Web que basan su estructura funcional en esquemas de metadatos y por ello han sido objeto de esta investigación.

Con todo, el desarrollo del cuerpo del trabajo, se ha estructurado en cuatro grandes apartados que se ocupan, a su vez, de cuatro aspectos bien delimitados: tres de discusión y debate científico en torno al complejo tema que nos ocupa (partes I, II y III) y un proyecto de aplicación práctica (parte IV):

- La **parte I** (*De la información a la metainformación: metadatos*) plantea un acercamiento al estado de la cuestión, incluyendo una reflexión en torno al concepto e implicaciones de los metadatos así como su fundamento —el documento entendido como Objeto de Información (DLO)—, aplicación y tipología (**capítulo 3**). También se aborda el problema de la asignación de metadatos comparándolo, en primer lugar, con la catalogación tradicional, analizando tanto los métodos de creación y almacenamiento de metainformación, como las herramientas y aplicaciones de software desarrolladas a tal efecto (**capítulo 4**). En el capítulo 4 se parte de que el fundamento epistemológico de los metadatos para los bibliotecarios subyace en la catalogación. Al surgir nuestra investigación en el seno de la Documentación es inevitable tener como referente el proceso técnico tradicional. No obstante por otra parte, se deja patente el cambio de paradigma que supone el contexto de la información electrónica y se plantean

del software vinculado a los metadatos y a los dos entornos informativos que reconocemos: la recuperación en toda la Web, por un lado, donde los metadatos hoy se reducen a la aplicación de metaetiquetas HTML, y las bibliotecas digitales por otro, donde los metadatos pueden responder a sistemas complejos de estructuración y recuperación de información.

cuestiones prácticas para la aplicación de metadatos, tanto desde el punto de vista del agente-"catalogador" como de las aplicaciones de software.

- La **parte II** (*Normalización y metadatos: formatos, esquemas o estándares*), trata de reflejar tres realidades de las que estamos siendo testigos en el desarrollo de metadatos: la proliferación de diferentes esquemas o modelos de metainformación, el intento de normalizar su uso y la pretensión de reconciliar los atributos comunes con los imbricados estándares existentes. Así, este apartado se compone de dos capítulos perfectamente diferenciados: el estudio de los principales formatos o modelos de metadatos, haciendo hincapié en los lenguajes de marcado (SGML/HTML/XML), defendiendo la estructuración de la información para la descripción, recuperación e intercambio de recursos electrónicos (**capítulo 5**); y la cuestión fundamental de la normalización en un entorno tecnológicamente cambiante, pero vinculado al mundo bibliotecario de gran tradición normativa, planteando la necesidad de estándares *de facto* a partir de especificaciones de dominio público, y haciendo un repaso a la formalización de normas de metadatos; finalmente también se apunta el tema de la normalización del vocabulario en el ámbito de los metadatos ya que, desde nuestro punto de vista, cualquier estudio relacionado con la búsqueda y recuperación de información o la evaluación de sistemas de información debe tener en cuenta la naturaleza interpretativa e insoslayable de la indización por materias, que abre además una nueva expectativa de investigación partiendo de los metadatos orientados al contenido (**capítulo 6**).

- En la **parte III** (*Acceso global a la información: recuperación, metadatos y bibliotecas digitales*), se relacionan los metadatos con su aplicación principal: la recuperación de información. Para ello se tratan dos vertientes de implementación de metadatos: por una parte, los sistemas de búsqueda en Internet de carácter general, es decir, la recuperación de información en la WWW (**capítulo 7**), y por otra, las bibliotecas digitales como contexto finito y abarcable donde los metadatos pueden alcanzar un nivel de control y de

operatividad real y concreto (**capítulo 8**). En el capítulo 7 se establece el contexto y las características básicas de la búsqueda de información en la Red, incidiendo en dos aspectos fundamentales: el concepto de *navegación* y el problema de la relevancia; se realiza además un análisis de las tendencias de recuperación en la Red haciendo hincapié en función de los metadatos, y se comparan las ventajas e inconvenientes de la búsqueda en texto completo frente la búsqueda basada en metadatos. En el capítulo 8 partimos de la biblioteca como concepto y reflexionamos en torno a las bibliotecas digitales/virtuales/electrónicas y a su organización basada en metadatos. Se observan asimismo, las implicaciones de la globalización de la información para las bibliotecas digitales, y cómo afecta el acceso universal al planteamiento organizativo de metadatos y a la interoperabilidad entre los distintos esquemas, que fundamentan a su vez diversos sistemas de información digital.

- La **parte IV** está constituida por un solo capítulo (**9**) donde se presenta un proyecto de adaptación de metadatos a la Biblioteca Virtual *Miguel de Cervantes*. A lo largo de este apartado se trata de armonizar, a través de una reflexión de aplicación práctica, toda la dialéctica abordada en los capítulos anteriores en torno a los metadatos, su competencia como herramienta para la recuperación de información y el planteamiento de aplicación en el marco de esta biblioteca digital paradigmática en el ámbito informativo en español.
- Además, se incluye un capítulo de conclusiones (**10**) donde se resumen la confrontación de hipótesis y los resultados a los que hemos llegado en esta tesis, aglutinando tanto consideraciones de orden teórico y/o conceptual, como aspectos experimentales e inferencias acerca del desarrollo de los metadatos o de su implantación.

Este es el desarrollo esquemático del contenido o del núcleo de la tesis que, por motivos metodológicos y para favorecer la fluidez en el tratamiento de los distintos aspectos reflejados, tiene la estructura que acabamos de presentar. La longitud de

cada uno de los capítulos, que podría parecer irregular en algunos casos, responde a la profundidad que se ha arbitrado a cada uno de los temas, con la finalidad de alcanzar los objetivos de esta investigación. Es importante señalar no obstante, que en el entramado discursivo de todo el trabajo, existen capítulos íntimamente relacionados que podría haber dado lugar a otra ordenación de los mismos, y aún a la inclusión de unos en otros. Por ello, hemos hecho constar, a lo largo de toda la redacción del trabajo, referencias cruzadas que remiten de unos apartados a otros con el fin de homogeneizar la tesis. Algunas de las relaciones más evidentes las resumimos a continuación:

- La clasificación de los metadatos en zonas que se hace en el capítulo 3 (3.5) estará presente también al hablar de los esquemas (*schemas*) o modelos de metainformación (capítulo 5).
- Las reflexiones en torno al formato MARC (4.1) se reflejarán también al hablar de la estructuración de la información (5.1.3.) y también bajo el epígrafe que aborda el problema de la normalización de metadatos (6.1).
- La evaluación de herramientas y aplicaciones para la gestión de metadatos que realizamos en el apartado 4.3, anticipa algunas de las ideas que se discutirán a lo largo de toda la tesis, tomándolas como elemento de juicio tácito para evaluar el software (*v. gr.* opciones de visualización en XML, la integración con otros módulos de software, o la posibilidad de trabajar con *schemes*¹³ o estándares de contenido).
- La clasificación de herramientas expresada en el capítulo 4 tiene la lógica implícita del planteamiento que se hace de la recuperación de información en la parte III de la tesis. Las aplicaciones informáticas destinadas a la creación de metaetiquetas (4.3.1.) tienen como finalidad asignar metainformación destinada a los sistemas de búsqueda genéricos de la Web (capítulo 7), cuyo

¹³ Vid. Glosario, *schema/scheme*.

modelo de propósito general responde fundamentalmente, a estructuras genéricas de metainformación, esto es, a etiquetas <META> de HTML (5.2.1.1.) o a lo sumo, a la semántica del Dublin Core (5.2.1.2.) embebida en ellas. Por el contrario, las aplicaciones para la asignación de metadatos (4.3.2) responden a las expectativas de sistemas de metadatos aplicados en entornos finitos de información o bibliotecas digitales (capítulo 8), donde los esquemas de metadatos son generalmente, de propósito específico (5.2.2) y tienden a una sintaxis, originalmente en SGML, y cada vez más en XML.

- El capítulo 6, donde se plantea y reflexiona acerca del papel de la estandarización en todo este amplio contexto tiene, por un lado su esencia o raíz en el capítulo 5 que analiza los formatos de metadatos y, por otro tiene la continuación lógica de los argumentos técnicos de la normalización, en el apartado 8.3.2 donde se habla de la interoperabilidad entre estándares de metainformación en el amplio contexto de las bibliotecas digitales. El tema de la interoperabilidad y del establecimiento de *crosswalks*¹⁴ se ha preferido tratar bajo el concepto del acceso universal ante la globalización de la información, por ser éste un marco más extensivo con repercusiones sociales; de esta forma podemos considerar los problemas que suscita la interoperabilidad con una perspectiva más amplia, no sólo a nivel técnico, sino también a nivel político.
- En el capítulo 9 confluyen todos los demás, no sólo los aspectos conceptuales sobre metadatos y bibliotecas digitales (capítulos 3 y 8, respectivamente), sino todas las cuestiones discutidas en la tesis, pero desde una óptica de aplicación real, concreta y prospectiva.

En cuanto al contenido del trabajo, tenemos que hacer también referencia al aparato crítico que, incluido a pie de página, resulta un componente esencial para el

¹⁴ Vid. Glosario.

desarrollo de la investigación. Se ha elaborado según el sistema cita-nota, donde las notas a pie de página no sólo reflejan la fuente consultada, sino que pretender ser además un aporte intelectual a este trabajo, bien a través del establecimiento de relaciones y referencias cruzadas, a las que aludíamos anteriormente, bien a través de digresiones que completan o tratan de enriquecer el resultado final del mismo, contrastando nuestras aportaciones con las fuentes.

La redacción de las referencias bibliográficas se ha realizado según la Norma ISO 690 (1987) - PNE 50-104¹⁵ para el caso de los documentos impresos; y para los documentos electrónicos y/o de Internet, a pesar de que existe un estándar ISO¹⁶ específico, se ha seguido un estilo personal de referencia¹⁷ basado en diferentes

¹⁵ Documentación. Referencias bibliográficas. Contenido, forma y estructura: ISO 690-1987, PNE 50-104. *Revista Española de Documentación Científica*, 1992, vol. 15, nº 3, p. 281-288; 1993, vol. 16, nº 1, p. 53-68.

¹⁶ *Excerpts from International Standard ISO 690-2 Information and Documentation: Bibliographic References. Part 2: Electronic Documents or Parts Thereof* [documento HTML]. ISO, 2000, rev. 14 de septiembre de 2000. Disponible en: <http://www.nlc-bnc.ca/iso/tc46sc9/standard/690-2e.htm> (consultado el 3 de noviembre de 2000).

¹⁷ Bonifacio Martín Galán, Eva M^a Méndez Rodríguez. *Cómo citar y referenciar documentos electrónicos. Particularidades e iniciativas*. Trabajo presentado en el Doctorado en Documentación. Universidad Carlos III de Madrid, 1997, p. 39-42, (ejemplar original de autor). A pesar de que nos basamos esencialmente en la norma ISO 690-2 (*Ibid.*) en lo referente a la información que es preciso consignar, el modelo de referencia que proponemos trata de caracterizar más puntualmente las particularidades de la información de Internet. Las únicas variaciones de nuestro formato de citación con respecto a dicha norma son:

- a) La referencia [online] del tipo de documento, nos parece extremadamente genérica y no anticipa ni identifica la casuística de los formatos de codificación electrónica que encontramos en la Red, y que por su importancia aparecen citados y referenciados en este trabajo (documentos HTML, documentos PDF, PS, RTF, presentaciones PPT, etc.). El hecho de consignar el tipo de documento o fichero informa al lector acerca del software necesario para visualizarlo sin alargar inútilmente la referencia.
- b) Por otra parte, la norma ISO (*Ibid.*, punto 7.9.4) determina la fecha de citación, que nosotros hemos cambiado por fecha de consulta, ya que en una investigación de estas características varía la fecha en que se consulta, se mira, se imprime, etc. una información de Internet, del momento en que se utiliza para la redacción del trabajo.
- c) En cuanto al formato de las fechas, decidimos ponerlo en orden directo (día mes año) para aumentar la facilidad de comprensión.

[cont.]

estilos propugnados para identificar los recursos electrónicos (ISO 690-2, MLA, APA, etc.). Nuestra forma de consignar los recursos de la Red trata de recoger el máximo número de datos que identifiquen la publicación electrónica: autor; título; lugar (si es seguro o aparece expresamente en el servidor donde está la fuente que citamos); institución científica, académica o comercial, en cuyo servidor se ha publicado electrónicamente el recurso; fecha de edición y/o revisión; el URL (*Uniform Resource Locator*) más exacto posible relativo al documento o fragmento del mismo que citamos, precedido de la frase "disponible en"; y finalmente, la fecha en la cuál hemos consultado tal documento entre paréntesis. Es importante señalar en este sentido que, dada la fluctuación habitual del contenido de los recursos Web, máxime si se trata de informaciones en estado de desarrollo e investigación, y por ello, constante evolución como la que aquí se trata, se ha consignado en nota al pie la fecha de consulta del recurso en el momento de acceso o de la redacción del trabajo. Sin embargo en la bibliografía final se han revisado todos los enlaces y las referencias a los recursos electrónicos, a fin de presentar una bibliografía general sobre el tema actualizada en el momento de depósito de este trabajo (Diciembre de 2001).

Dentro del estilo de citación, es preciso comentar que en las notas a pie de página se incluye, la primera vez que se cita una obra, la referencia íntegra de la misma (sin inversión del nombre y apellidos del autor¹⁸), sin que obste para incluir una bibliografía final completa, actualizada y alfabética por autores. En el caso de que una misma obra aparezca citada en distintas partes de la tesis, no se consigna la

d) Finalmente, añadimos en algunos casos el elemento *fecha de revisión*, precedido de la abreviatura "rev." para consignar la última fecha de actualización y revisión del documento en el momento de consulta. En el caso de que no exista esa fecha, hemos recurrido a la opción del navegador Netscape *document info* ó *información del documento* para ver la fecha en que el documento citado fue transferido por última vez al servidor que lo alberga. Es los casos en que esta fecha está oculta, hemos recurrido también a la herramienta de generación de metadatos DC-Dot, que siempre extrae como fecha del documento la fecha de transferencia del fichero, esto es, la fecha efectiva de publicación en la Web. *Cfr.* Capítulo 4, nota 143.

¹⁸ Según indica el punto 9.3.1. de la Norma. Documentación. Referencias bibliográficas. *Op. cit.*, 1993, vol. 16, nº 1, p. 66.

referencia completa pero sí se alude a ella a través de las abreviaturas latinas de uso frecuente en investigaciones del área de humanidades y ciencias sociales, principalmente: *Op. cit.* (obra citada del autor en una nota previa), *Ibid.* (mismo autor y obra que en la nota inmediatamente anterior), etc. precisando la página concreta o el URI más preciso donde aparece la cita. En las notas al pie, igualmente se utilizan abreviaturas de relación del tipo: *Cfr.* (confrontar o comparar con otro apartado del trabajo o con un autor u obra concretos), *Vid.* (véase, remitiendo a otro lugar de la investigación o a un autor que trata o amplía el tema que se cita).

La importancia de la terminología

Las disciplinas en desarrollo, como la Documentación, deben luchar con las palabras para encontrar el término justo para designar los avances y los nuevos conceptos que surgen a tenor de ellos. Internet ha provocado efectos curiosos sobre la lengua y ello ha llevado a numerosas reflexiones¹⁹ acerca de las amenazas que acechan a nuestro idioma por crecimiento terminológico que surge a raíz de la Red. Así, en el último *Congreso Internacional de la Lengua española*²⁰ se incluyeron como cuestiones fundamentales: el español en la Sociedad de la Información o el español en la ciencia y la técnica al albur de las nuevas tecnologías, la presencia de

¹⁹ Son muchos los autores que reconocen este influjo de la Red de redes. Millán, por ejemplo, consigna el paralelismo de crecimiento lingüístico provocado por Internet con lo que ocurriera en épocas pasadas con dominios como el deporte o la técnica. José Antonio Millán. *Internet y el español*. Madrid: Fundación Retevisión, 2001, p. 175 y ss. El resultado de la influencia de las TIC y de Internet en el español ha sido una gran proliferación de términos ingleses con muy distinto grado de asimilación, que han dado lugar, incluso, a lo que se denomina *spanGLISH* (términos ingleses en español, que suelen ser barbarismos en nuestra lengua como v. gr. *resetear*, *forwardear*, etc.).

Ese otro "idioma" o jerga denominada *cyberspanGLISH* o ciberespanglis, surgido por la gran cantidad de términos y palabras introducidas en nuestra lengua, se ha reflejado en glosarios como el de la Universidad de Texas <<http://www.actlab.utexas.edu/~seagull>> presentado en 1998. Yolanda Rivas. "Surfea" en el cyberspanGLISH [documento HTML]. *El Mundo*, viernes 19 de junio de 1998. Disponible en: <http://www.el-mundo.es/navegante/diario/98/junio/19/cyberspanGLISH.html> (consultado el 3 de marzo de 1999).

²⁰ *II Congreso Internacional de la Lengua Española*, celebrado en Valladolid del 15 al 19 de octubre de 2001. *Vid.* Sitio Web de dicho congreso: <http://congresodelalengua.cervantes.es>

nuestro idioma en la Red y las mutaciones que experimenta en y por ella. De igual forma, en esta investigación hemos querido dar cuenta de esta cuestión fundamental, de la influencia que ejerce la creación científico-técnica en las fronteras conceptuales de las lenguas. Por ello, incluimos un glosario explicativo para mejorar la comprensión final del trabajo y justificar el uso que hacemos de la nueva terminología asociada a las bibliotecas digitales y los estándares de metainformación.

La repercusión de Internet sobre el español es inminente no sólo en nuestro lenguaje científico-técnico, sino también en el de otras disciplinas e incluso a distintos niveles comunicativos de la sociedad. Esto se debe a que el vocabulario base de Internet proviene del inglés, y su inserción en la lengua provoca tensiones y confusiones que pueden entorpecer el avance de la investigación en nuestro idioma. La preocupación por la lengua ha alcanzado a colectivos que antes permanecían ajenos al debate lingüístico, como por ejemplo los informáticos, quienes en muchos casos, se han preocupado por reflexionar sobre equivalentes españoles, definiciones, etc. para la creciente terminología inglesa aparejada a la investigación informática²¹.

Esta misma preocupación se ha reflejado en el contexto profesional de bibliotecarios y documentalistas donde, aunque de forma menos recurrente o sistemática, han aparecido algunas herramientas terminológicas y/o lexicográficas para salvar la barrera lingüística del inglés en el desarrollo de los estudios y de la investigación de nuestra disciplina²².

²¹ En este sentido, debemos destacar la labor de algunos informáticos como: Antonio Vaquero, Rafael Fernández Calvo, M^a del Carmen Ugarte, etc. y, en general, del grupo de trabajo surgido, por este motivo, en el seno de la Asociación de Técnicos de Informática (ATI): <http://www.ati.es/gt/lengua-informatica>. Asimismo han aparecido diccionarios y glosarios como el de Rafael Fernández Calvo. *Glosario básico inglés-español para usuarios de Internet*. 4^a ed. Con vocabulario español-inglés. Barcelona: ATI, 2001. También disponible en Internet en: <http://www.ati.es/novatica/glointv2.html>

²² Por ejemplo: Antonio Lozano Palacios. *Vocabulario inglés/español – español/inglés para los estudios de biblio-documentación*. 3^a ed. rev. y aum. Granada: Universidad de Granada, Facultad de Biblioteconomía y Documentación, octubre 1996. La actualización Web de este diccionario, enriquecido con la búsqueda en otros glosarios específicos de Biblioteconomía y Documentación (como el de la IFLA) se puede consultar en Internet <<http://eubd1.ugr.es/RIS/risWeb.isa>>, donde
[cont.]

El crecimiento lingüístico que ha experimentado el área de conocimiento de Biblioteconomía y Documentación en los últimos años se debe fundamentalmente a Internet y a la nueva gestión de la información electrónica, donde han surgido múltiples neologismos —que en muchos casos son cultismos, como el propio término metadatos (*metadata*²³), barbarismos, como p. ej. nestado (*nesting*) o metáforas como *crosswalk* o *gateway* o, en un ámbito de aceptación generalizado, portales— para denominar nuevas realidades técnicas en la organización y gestión de la información digital, propiciada y demandada por la explosión informativa de la World Wide Web. A la complejidad del crecimiento semántico se une el problema de la traducción de muchos de estos términos al español. Estos problemas de traducción, concretamente de los nuevos términos relacionados con los metadatos, se han reconocido en nuestro ámbito profesional el último congreso de la IFLA²⁴. No obstante, la integridad de un idioma depende más de cómo incorpora y desactiva las invasiones externas y no tanto de cómo las elimina o excluye.

Por estas circunstancias y tomando en cuenta que más del 90% de la literatura significativa consultada para esta tesis se encuentra en lengua inglesa, hemos considerado oportuno incluir un glosario con el propósito de que contribuya a facilitar la comprensión de esa terminología científica y su traducción a nuestra lengua. No aspira a ser una herramienta lexicográfica constituida formalmente, ni

además del significado o traducción de cada uno de los términos, se adjunta un ejemplo en inglés de las palabras en el contexto del significado al que se refieren. Tras su última actualización —febrero de 2001— dicho diccionario electrónico contaba con 49.269 entradas.

²³ Cfr. 3.1. donde se presentan distintas reflexiones sobre la etimología y significado del término fundamental en esta tesis doctoral.

²⁴ Incluso en el ámbito de la IFLA, cuya posición en torno a los metadatos sería la de un enfoque eminentemente bibliográfico, tendente a simplificar los metadatos a la catalogación (Cfr. Burnett, Bor Ng y Park, 3.2.), se reconoce expresamente el crecimiento terminológico a tenor de la adopción de formatos como el Dublin Core en las bibliotecas. Así Muennich dice literalmente: *The Dublin Core Metadata scheme introduced many new "terms" that now require "official" translations, too.* Monika Muennich. Progress on the Multilingual Dictionary of Cataloging Terms and Concepts [documento HTML]. En: *IFLA Council and General Conference (67. 2001. Boston)* rev. 11 de noviembre de 2001. Disponible en: <http://www.ifla.org/IV/ifla67/papers/093-152ae.pdf> (consultado el 20 de noviembre de 2001).

pretende extraer los conceptos más importantes utilizados a lo largo del mismo, como si se tratase de un manual. Si bien es cierto que esta investigación daría lugar a un amplio glosario terminológico más convencional, el que hemos elaborado es, más bien, un breve comentario en torno al concepto, traducción y uso en español de algunos vocablos que aparecen constantemente en la bibliografía anglosajona relacionada con los metadatos y las bibliotecas digitales y, por ende, en este trabajo. En cierto modo, se trata también de un pequeño "diccionario de autoridades" pues, en la mayoría de los casos, cada significado y/o utilización de los términos está autorizado por un testimonio, no literario pero sí científico, de especialistas en la materia seleccionados de entre la bibliografía de este trabajo, para justificar el trasvase o el uso que hacemos de cada expresión surgida originalmente en el dominio lingüístico anglosajón.

De esta forma, las matizaciones y explicaciones que incluimos para cada vocablo o locución en el glosario sirven para justificar, bien el uso directamente en inglés, su traducción (a través del calco lingüístico o semántico), o la españolización de algunas de las palabras que aparecen en la tesis, procurando con ello mejorar su comprensión²⁵. La herramienta terminológica que creamos fruto de esta investigación, sirve además para poner de manifiesto que, a partir de su complejo crecimiento terminológico, la Biblioteconomía y Documentación es una disciplina en eferescencia, vinculada estrechamente al avance tecnológico e informático, así como para demostrar el vacío lingüístico en español²⁶ de muchos de los conceptos

²⁵ Por ello, consideramos el Glosario una parte fundamental de este trabajo y serán frecuentes las remisiones al mismo a través de notas al pie.

²⁶ Antonio Vaquero comenta que en el I Encuentro de la Lengua Española, celebrado en Zacatecas (México) en abril de 1997 ya se dijo, entre muchas otras cosas, que una de las partes más débiles del español es la tecnológica, a lo que argumentaba que *independientemente de lo que se entienda por débil, la comunidad científica debe hacer un examen de conciencia y quizá un propósito de enmienda*. Antonio Vaquero. La corrupción del lenguaje por la informática [documento HTML]. *El Mundo*, 17 de abril de 1997. Disponible en: <http://www.el-mundo.es/1997/04/17/opinion/17N0017.html> (consultado el 19 de abril de 2001); y Antonio Vaquero. La lengua española en el contexto informático [documento HTML]. *Novática*, Julio-agosto 1999, n° 140. Disponible en: <http://www.ati.es/novatica/1999/140/antvaq140.html> (consultado el 12 de mayo de 2001). Según estas palabras del [cont.]

surgidos a tenor de los temas más actuales surgidos en su seno, como es el de los metadatos.

A pesar de todo, no aparecen en el glosario la totalidad de las palabras inglesas que utilizamos a lo largo de este trabajo, bien porque son ya de uso común (*software, byte, applet, homepage, script, etc.*) “universalmente aceptadas”, o porque no requieren una explicación adicional. Sí recogemos todos aquellos términos para los cuales hemos arbitrado una traducción propia, o los que hemos utilizado directamente en la lengua original que designan un concepto complejo (como *spamming*) o un marco semántico concreto (*v. gr. harvesting*).

Siglas y acrónimos

Una de las manifestaciones más evidentes de la nueva cultura digital y de la nueva Sociedad de la Información, marco estructural de esta tesis, es la *jerga que altera el lenguaje habitual en forma de siglas*²⁷. La Documentación, que en sí misma ya cuenta con gran número de siglas y acrónimos (*v.gr. ISBD, RC, AACR, etc.*), ha aumentado esta forma de reducción sintáctica y de simplificación designativa a raíz de la tecnología y de Internet. Así, en el ámbito concreto de la metainformación y de las bibliotecas digitales, dominio terminológico que hemos manejado aquí, proliferan siglas para denominar tanto servicios de información, como modelos de metadatos y proyectos e instituciones relacionados con ellos.

En este trabajo se utilizan pues, profusamente las siglas y los acrónimos. En algunos casos, porque tienen una aceptación tan generalizada que resulta ser más

profesor Vaquero, el glosario que incluimos en esta tesis es parte de ese “examen de conciencia” que debemos hacer los investigadores españoles, adoptando el compromiso de tratar de encontrar términos en nuestra lengua para denominar los nuevos conceptos que surgen de forma vertiginosa a tenor de la tecnología, o al menos, de justificar que nuestros trabajos de investigación estén teñidos de términos en inglés, al mismo tiempo que aclaramos su significado y carga semántica.

²⁷ José B. Terceiro. Prólogo. En: José B. Terceiro y Gustavo Matías. *Digitalismo: el nuevo horizonte sociocultural*. Madrid: Taurus, 2001, p. 14.

conocido el acrónimo o la sigla correspondiente que su significado completo (p. ej. MARC, DC, TEI, RDF, etc.); en otros, hemos constituido siglas *ad hoc* para abreviar algunos conceptos (p. ej. SRII, Sistemas de Recuperación de Información en Internet o BVC, Biblioteca Virtual Cervantes). Aunque estas siglas o acrónimos están explicados convenientemente en sus respectivos contextos (bien a través del propio discurso expositivo, bien en notas al pie de página), nos ha parecido conveniente reunirlos en un índice final que contiene aquellos relacionados con la temática de esta tesis, acompañado de su significado y, en su caso, del URL de la ubicación en la World-Wide Web del recurso/s relacionados con el tema o institución que evocan dichas siglas, de tal forma que pueda servir también como referente bibliográfico sobre los modelos, instituciones o/y sistemas citados²⁸.

2.3. Fuentes y bibliografía utilizada

Como ya hemos mencionado, los metadatos y las bibliotecas digitales constituyen un área aún emergente dentro de la investigación en representación y recuperación de la información, surgida, o al menos fomentada, a raíz de la universalización del fenómeno Internet. Por todo ello, no existe gran cobertura y actualización sobre el tema en la literatura impresa. A pesar de que las publicaciones sobre metadatos y bibliotecas digitales han ido *in crescendo* a lo largo del tiempo que hemos dedicado a este tema, como demostraremos a continuación, existen todavía pocos libros²⁹, y en los artículos de publicaciones periódicas, salvo algunas

²⁸ Al igual que en la bibliografía, se ha actualizado, a través del acceso y comprobación al final de la investigación (concretamente, en este caso, en la fecha 27 de noviembre de 2001), cada uno de los URLs consignados. *Cfr.* Nota inicial de la Bibliografía.

²⁹ En el transcurso de elaboración de esta investigación, han comenzado a publicarse varias monografías sobre metadatos y bibliotecas digitales, todas a partir del año 1998. El primero de estos libros, editado en Los Ángeles por Murtha Baca, en el contexto institucional del J. Paul Getty Trust: *Introduction to Metadata: Pathways to Digital Information*, se tradujo al español en 1999 con el título: *Introducción a los metadatos: vías a la información digital*. Murtha Baca, ed. Los Angeles: J. Paul Getty Trust, 1999, 44 p. Una edición aumentada, revisada y actualizada de este pequeño manual [cont.]

excepciones recientes, son poco significativos los tratamientos monográficos del tema. Esta carestía de publicaciones impresas se justifica obviamente por la obsolescencia que apremia a todos los asuntos de carácter tecnológico y fundamentalmente por el componente de innovación constante que subyace en la propia Red. Sin embargo, la actualización y la abundancia de recursos informativos en la Web sobre este tema, compensa, incluso en exceso, la escasa literatura impresa.

La primera publicación periódica en formato "tradicional" de la que tenemos constancia, que dedicó un número a este tema en toda la dimensión del problema que abarcamos aquí, fue la conocida revista *Bulletin of the American Society for Information Science* que, en octubre-noviembre de 1997, recogió una serie de artículos relacionados con los metadatos bajo el título *Organizing Internet Resources: Metadata and the Web*³⁰. A estas excepciones de la literatura impresa

introdutorio al tema de los metadatos, se publicó únicamente en inglés y en versión electrónica en la Web, el 1 de agosto del 2000 en: <http://www.getty.edu/gri/standard/intrometadata>, que posteriormente (el 13 de septiembre de 2001) cambiaría el URL de publicación, aunque no los contenidos, a: <http://www.getty.edu/research/institute/standards/intrometadata>. Este ejemplo demuestra que, dada la vertiginosa evolución del tema, la mejor forma de difundir los avances en él es la inmediatez editorial de la Web. No obstante, podemos destacar algunas monografías impresas de los últimos tres años:

- Jean Hudgins, Grace Agnew, Elizabeth Brown. *Getting Mileage out of Metadata: Applications for the Library*. Chicago: ALA, 1999, 79 p.
- Virpul Kashyap, Amit Sheth. *Information Brokering Across Heterogeneous Digital Data: A Metadata-based Approach*. Boston, Dordrecht, London: Kluwer Academic Publishers, 2000, 224 p.
- Johan Hjelm. *Creating the Semantic Web with RDF: Professional Developer's Guide*. New York, etc.: Wiley Computer Publishing, John Wiley & Sons, 2001, 277 p.
- Kal Ahmed, et al. *Professional XML Meta Data*. Birmingham: Wrox Press, 2001, 566 p.
- En español, aparte de la traducción del libro editado por Murtha Baca sólo existe otra publicación surgida en el ámbito latinoamericano: *Internet, metadatos y acceso a la información en bibliotecas y redes en la Era Electrónica*. Filiberto Felipe Martínez Arellano, Lina Escalona Ríos, comp. México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas; Infoconsultores, 2000, 112 p. Esta obra es a su vez una traducción de las actas del XV Coloquio de investigación del mismo título, celebrado en México los días 5, 6 y 7 de agosto de 1998. Más información sobre dicho congreso en: <http://cuib.unam.mx/cursos/infoesp.html>

³⁰ *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, nº 1. También es cierto, que esta publicación está en Internet en texto completo desde el vol. 21, nº 2, correspondiente a diciembre-enero 1995. Por ello, podríamos

[cont.]

pertenecen también los números monográficos que prestigiosas revistas de nuestro ámbito profesional dedicaron al tema en 1999 —año en el que, desde nuestro punto de vista, ya se puede hablar de una cierta madurez de los metadatos—. Destacan pues en ese año, el *Journal of the American Society for Information Science*, vol. 50, n° 13 (*Special Topic Issue: Integrating Multiple Overlapping Metadata Standards*), o los números 116 y 117 de la Revista *Vine (Very Informal Newsletter)*, que se dedicaron también por completo a este tema. Por otra parte, la revista *Journal of Documentation*, a pesar de no haber dedicado ningún número monográfico, es habitual que conceda un espacio en sus páginas a la investigación en metadatos y bibliotecas digitales.

El interés por los temas que trata esta tesis doctoral ha aumentado considerablemente en los últimos años. Los resultados de las búsquedas en las principales bases de datos (Inspec y LISA, *Vid. Tablas 1-4*³¹), por periodos temporales de cinco años³² en los últimos veinte, ante la consulta "*metadata*" y "*digital libr**",

considerar la versión electrónica del monográfico dedicado a los metadatos <<http://www.asis.org/Bulletin/Oct-97/index.html>> como una publicación electrónica más.

³¹ Los datos de estas tablas fueron tomados en la fecha 03/07/2001. La cobertura que presentan dichas bases de datos en esa fecha de consulta es:

- En el caso de LISA, de 1969 a junio de 2001.
- En el caso de Inspec de 1980 a abril del 2001 (esta base de datos está seccionada por periodos temporales, lo que nos ha permitido elegir el ámbito de la búsqueda excluyendo *a priori* el espacio temporal de 1969-1979 que no consideramos en esta muestra).

La cobertura particular de estas bases de datos justifica que, en el caso de LISA, los resultados que aparecen como totales, en las diferentes tablas que presentamos, se refieren al número de publicaciones que tienen la palabra buscada en toda la base de datos, y no la suma de las ocurrencias que aparecen en los años analizados. Así, p. ej., la suma de las publicaciones que contienen el término *metadata* en el campo título de la base de datos LISA en los periodos observados es 186, mientras que el resultado que consignamos como total en la Tabla 3 es el que da la búsqueda en ese campo en toda la base de datos sin acotar temporalmente (191); lo que implica que existen 11 publicaciones anteriores a 1980 que tienen dicho término en el título. En el caso de Inspec, al poder seleccionar *a priori* los años de búsqueda, los totales que aparecen en la columna "total" son los mismos que la suma de los resultados que aparecen por años.

³² Salvo en el último periodo (1996-2001) que se consignan aproximadamente cinco años y medio (exactamente cinco años y seis meses en LISA y cinco años y cuatro meses en Inspec), al estar realizadas las búsquedas en julio de 2001.

demuestran el crecimiento del interés científico, además de la proliferación de artículos en las principales revistas profesionales tanto del ámbito de la Documentación como de la Informática y la Tecnología. La tabla 1, refleja la frecuencia de aparición de los términos "metadata" y "digital libr*" (library/libraries/librarian, etc.) buscados en todos los campos de estas bases de datos. La tabla 2 destaca la bibliografía publicada en los cuatro/seis primeros meses del año 2001 (Inspec/LISA). La tabla 3, muestra los resultados de las mismas búsquedas en el campo de título (TI). Finalmente, la tabla 4, refleja los resultados en el campo dedicado al resumen (AB).

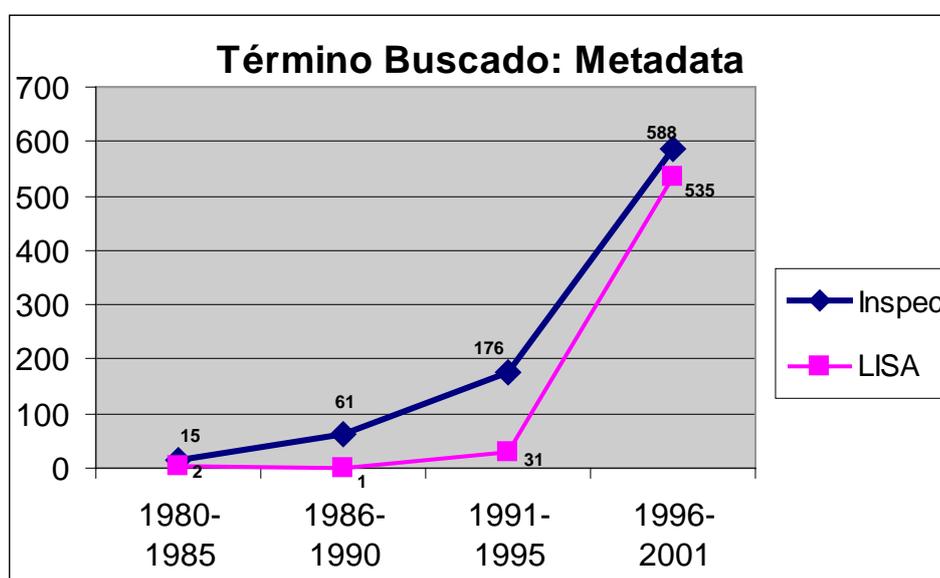
	Base De datos	1980-1985	1986-1990	1991-1995	1996-2001	Total
<i>Metadata</i>	Inspec	15	61	176	588	833
	LISA	2	1	31	535	605
<i>Digital libr*</i>	Inspec	7	18	188	1334	1551
	LISA	0	2	98	1122	1268

Tabla 1. Bases de datos: Resultados en todos los campos

Según evidencia la tabla 1, el incremento de las publicaciones en los últimos cinco años y medio es muy alto: del total de publicaciones sobre metadatos que aparecen en la base de datos Inspec (833), sólo el 9,1% pertenecen a los años ochenta; mientras que el 70,5% de los artículos se han publicado en los últimos cinco años y medio (1996-2001). Los datos son aún más evidentes en el caso de la base de datos LISA, donde el 88,4% de los artículos reflejados han visto la luz en el último periodo temporal analizado (1996-2001) y tan sólo el 0,49% de las publicaciones pertenece a la década de 1980-1990³³.

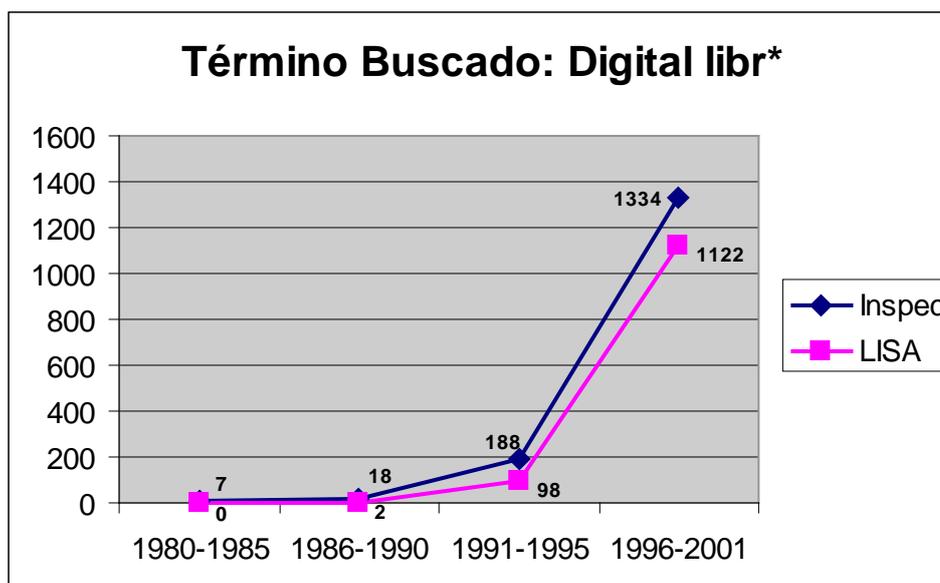
³³ Esta diferencia tan marcada entre una base de datos y otra en las publicaciones de los años 80 (9,1% en Inspec y 0,4% en LISA) se debe probablemente a que, como explicaremos (Cfr. 3.1), en el ámbito informático, el concepto de metadatos estaba vinculado en un principio a las bases de datos, mientras [cont.]

En la búsqueda por "biblioteca(s) digital(es)" los datos son semejantes: entre 1980 y 1990 se publicaron un 1,6% de los artículos que constan en la base de datos Inspec y un 0,15% de los que recoge LISA; mientras que de 1996 a abril del 2001, en el caso de Inspec, y a junio de este mismo año, en el caso de LISA, se publicaron el 86% y el 88,4%, de los artículos que recogen, respectivamente, dichas bases de datos bibliográficas. La progresión de este aumento de las publicaciones surgidas sobre los temas principales que abordamos en este trabajo se pueden ver los gráficos 1 y 2.



Graf. 1. Aumento de las publicaciones sobre "metadatos" en LISA e Inspec (1980-2001)

que en esa época, en el ámbito de la Biblioteconomía y la Documentación —área científica de las revistas que incluye LISA en su base de datos—nadie denominaba "metadatos" a los resultados del análisis documental.



Graf. 2. Aumento de las publicaciones sobre "bibliotecas digitales" en LISA e Inspec (1980-2001)

El crecimiento del interés por estos temas, que se perciben en los resultados que revelan estas bases de datos, es incluso sorprendente si analizamos las publicaciones que han aparecido en los últimos meses contemplados del año 2001 (tabla 2):

	Base De datos	Total	1996-2001	1996-2000	2001
<i>Metadata</i>	Inspec	833	588	561	27
	LISA	605	535	454	81
<i>Digital libr*</i>	Inspec	1551	1334	1304	30
	LISA	1268	1122	1019	103

Tabla 2. Publicaciones en el 2001 (Inspec-Abril / Lisa-Junio)

Según refleja la tabla 2, en el caso de la base de datos Inspec, las 27 publicaciones sobre metadatos aparecidas en sólo cuatro meses del año 2001 (enero-abril) representan un 3,2% del total de publicaciones que recoge esta base de datos sobre este tema, y el 4,5% de las aparecidas en el último tramo temporal observado

(1996-2001). En el caso de las publicaciones sobre biblioteca(s) digital(es), en esta misma base de datos, el 1,9% de las publicaciones totales sobre tal concepto, y el 2,2% de las publicaciones aparecidas entre 1996 y 2001 vieron la luz en los cuatro primeros meses del año 2001.

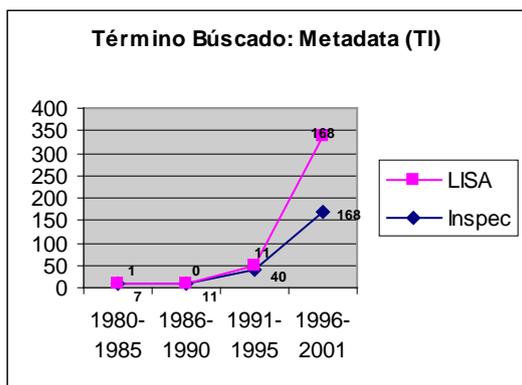
En el caso de LISA, los datos demuestran en mayor medida el aumento del interés científico internacional en el campo de la Documentación por los temas que trata esta tesis: el 13,3% del total de las publicaciones sobre metadatos que acopia esta base de datos y el 15,1% de las publicadas entre 1996 y 2001, se han publicado en los primeros seis meses del año 2001. De los artículos sobre biblioteca(s) digital(es), el 8,1% de todas las publicaciones sobre ese tema y el 9,1% de las aparecidas en los últimos cinco años y medio, se editaron de enero a junio del primer año del siglo XXI.

Si utilizamos la opción que permiten ambas bases de datos de acotar nuestras búsquedas a un campo de las mismas —título (TI) y resumen (AB)—, el crecimiento exponencial de las publicaciones es parecido, como demuestran las tablas 3 y 4 y las representaciones gráficas 3-4 y 5-6, respectivamente.

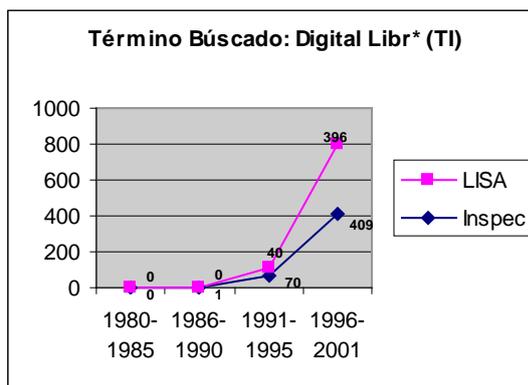
	Base De datos	1980-1985	1986-1990	1991-1995	1996-2001	Total ³⁴
<i>Metadatos</i>	Inspec	7	11	40	168	226
	LISA	1	0	11	168	191
<i>Digital libr*</i>	Inspec	0	1	70	409	481
	LISA	0	0	40	396	450

Tabla 3. Bases de datos: resultados en el campo de título (TI).

³⁴ Como comentamos en la nota 31 (*Cfr. sup.*), los totales relativos a la base de datos LISA no son la suma de las ocurrencias aparecidas en cada periodo temporal observado, sino el que da la base de datos sin acotar temporalmente (1969-2001).



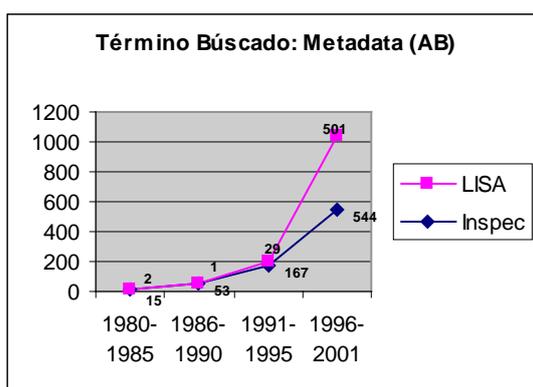
Graf. 3. Aumento de las publicaciones sobre "metadatos" en el campo título de LISA e Inspec (1980-2001)



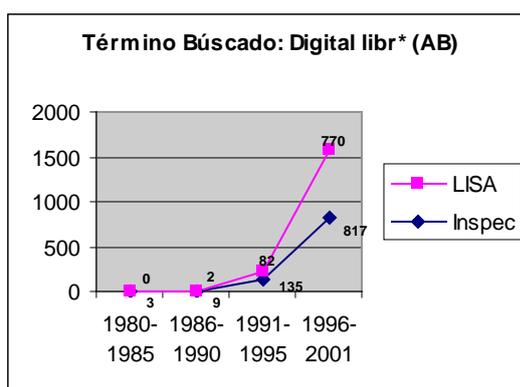
Graf. 4. Aumento de las publicaciones sobre "bibliotecas digitales" en el campo título de LISA e Inspec (1980-2001)

	Base De datos	1980-1985	1986-1990	1991-1995	1996-2001	Total
<i>Metadata</i>	Inspec	15	53	167	544	779
	LISA	2	1	29	501	563
<i>Digital libr**</i>	Inspec	3	9	135	817	967
	LISA	0	2	82	770	892

Tabla 4. Bases de datos: resultados en el campo de resumen (AB).



Graf. 5. Aumento de las publicaciones sobre "metadatos" en el campo resumen de LISA e Inspec (1980-2001)



Graf. 6. Aumento de las publicaciones sobre "bibliotecas digitales" en el campo resumen de LISA e Inspec (1980-2001)

Este crecimiento del interés por el tema, es aún más evidente si analizamos las publicaciones electrónicas en la WWW, tomando para ello los resultados de los dos motores de búsqueda más potentes³⁵:

- Altavista, en su opción avanzada admite realizar consultas por rango de fecha³⁶ <<http://www.altavista.com/sites/search/adv>> lo que nos permitirá soslayar —no con la misma certeza que las bases de datos bibliográficas analizadas anteriormente— el crecimiento de las publicaciones Web a través del tiempo.
- Google <<http://www.google.com>>. Aunque no permite hacer consultas limitadas temporalmente sí nos indica la importancia, desde el punto de vista cuantitativo, de los temas tratados.

³⁵ Se han seleccionado estos buscadores y no otros para el muestreo porque son, según los servicios de evaluación de motores de búsqueda en Internet (Searchenginewatch y Searchengineshowdown), respectivamente: el más exhaustivo en cuanto a la cobertura de su base de datos de información Web, en el caso de Google, <<http://www.searchengineshowdown.com/features/google/review.html>>, y por las potentes características de la búsqueda, en algunos casos exclusiva (como la búsqueda temporal) en el caso de Altavista <<http://www.searchengineshowdown.com/features/av/review.html>>. No obstante, y como trataremos en el capítulo 7 de este trabajo, tenemos que reconocer las limitaciones que tienen estos sistemas de recuperación de información en cuanto a la precisión y relevancia de la búsqueda en texto completo, además de la fluctuación constante de resultados, y estimar los datos que reflejamos en las tablas 5 y 6, con un valor relativo o tentativo.

³⁶ Según las opciones que ofrece Altavista para la consulta de fecha y teniendo en cuenta los períodos analizados en la consulta a bases de datos (de cinco en cinco años en los últimos 15), los rangos de referencia para este análisis de la bibliografía Web han sido: 01/01/80-31/12/85, 01/01/86-31/12/90, 01/01/91-31/12/95, 01/01/96-22/06/01.

	Buscador	1980-1985	1986-1990	1991-1995	1996-2001
<i>Metadata</i>	Altavista	3	1	811	592.072
	Google	880.000			
<i>Digital libr*</i>	Altavista	8	3	1.227	715.029
	Google	566.000			

Tabla 5. Bibliografía en la Web (en inglés)³⁸

Nuestra metodología fue realizar este muestreo en un mismo momento (22/06/2001) aprovechando la opción de Altavista de establecer escalas de fecha para las búsquedas (lo cual no es del todo exacto ni fiable, dada la versatilidad de las fechas de los documentos electrónicos). Otros autores reflejan estudios parciales semejantes a este para reflejar el aumento del interés por el tema en el tiempo, utilizado una metodología de observación diferente, por ejemplo, haciendo búsquedas totales sobre la base de datos de Altavista, en dos momentos diferentes. Así, Harter³⁹ se refiere al aumento de las publicaciones en Internet sobre bibliotecas digitales a través de los datos obtenidos, realizando la misma búsqueda en Altavista ("digital library" OR "digital libraries") en un lapso de seis meses. En Julio de 1996, dicha consulta en el popular buscador produjo 20,000 resultados, seis meses después (enero 1997), la misma búsqueda originó 30.000 ocurrencias, que interpreta como un aumento significativo [33%] de la importancia e interés del tema en sólo medio año.

³⁷ Recordemos que Google no permite el truncamiento para sus búsquedas.

³⁸ Datos tomados en la fecha: 22/06/01

³⁹ Stephen P. Harter. Scholarly Communication and the Digital Library: Problems and Issues [documento HTML]. *Journal of Digital Information*, 1997, vol. 1, issue 1. Disponible en: <http://jodi.ecs.soton.ac.uk/Articles/v01/i01/Harter> (consultado el 24 de julio de 1999).

	Buscador	1980-1985	1986-1990	1991-1995	1996-2001
Metadatos	Altavista	0	0	0	3.072
	Google	98.460			
Bibliotec* digital*	Altavista	0	0	9	9.900
	Google	15.400			

Tabla 6. Bibliografía en la Web (en español)⁴⁰

La tabla 6 refleja los resultados para las mismas búsquedas en español. Los datos reflejados en ella muestran que la bibliografía Web sobre estos temas en nuestra lengua es mucho más escasa que en inglés. Además, a través de la búsqueda por periodos que permite Altavista, podemos ver que se ha gestado casi en su totalidad —salvo los 9 documentos encontrados de 1991 a 1995 sobre biblioteca(s) digital(es)— en los últimos cinco años. La lectura y análisis de muchas de estas publicaciones a lo largo de esta investigación, nos ha demostrado además, que la profundidad y el nivel de especialización de las publicaciones en español, es inferior al de las del dominio lingüístico anglosajón.

Con todo, a pesar de que la información que consignamos a este respecto es meramente cuantitativa, lo que sí podemos afirmar es que la mayoría de las mejores fuentes sobre metadatos y/o bibliotecas digitales están publicadas en la Web. Destacamos, desde un punto de vista cualitativo ahora, su actualidad, tanto en el caso de revistas electrónicas, actas de congresos y seminarios que se celebran con cierta periodicidad (como los el *Institute of Electrical and Electronics Engineers*, los talleres del *Dublin Core* o las conferencias del W3C sobre la WWW), como en los que se efectúan de forma ocasional (v. gr. *Bicentennial Conference on Bibliographic Control for the New Millennium*), así como proyectos europeos (como DESIRE,

⁴⁰ Datos tomados en la fecha: 25/06/01

ROADS, BIBLINK o Renardus) o simplemente contribuciones particulares⁴¹ en formato HTML o XHTML y accesibles vía Web.

De esta forma se justifica que la mayor parte de las fuentes que, de manera efectiva, hemos utilizado para la realización de esta tesis doctoral sean publicaciones en formato electrónico y distribuidas en Internet, normalmente en inglés. Podemos destacar, a nivel general, por la cobertura que realizan del tema, la revista electrónica *D-Lib* <<http://www.dlib.org>> que se centra en todos los aspectos relativos a las bibliotecas electrónicas o virtuales, incluyendo varios artículos relativos a los metadatos, y la revista *Ariadne* —cuya versión electrónica está disponible en: <http://www.ariadne.ac.uk>—, que recoge todos los avances del Programa de Bibliotecas Electrónicas (eLib) y describe y evalúa los servicios y recursos accesibles en Internet que pueden ser de utilidad para los profesionales interesados en esta materia. Otras publicaciones electrónicas como el *Journal of Digital Information* (JoDI) o el *Journal of Electronic Publishing* (JEP) o boletines como el *RLG DigiNews* o *The National Digital Library Program Reports* de la biblioteca del Congreso de los Estados Unidos, o *Cultivate Interactive* o *eCulture*, que recogen los avances de la política cultural digital de la UE, han aportado también, informaciones puntuales interesantes.

Es importante subrayar que distintas instituciones mantienen en la Web recursos ampliamente detallados y actualizados sobre estos temas. Tal es el caso de organizaciones profesionales internacionales como la IFLA (*International Federation of Library Associations*), que alberga en sus servidores fuentes con enlaces a

⁴¹ A este tipo de aportaciones particulares pertenecen, por ejemplo, los documentos de Cathro o Hodson, publicados en los servidores de las instituciones donde trabajan respectivamente. Warwick Cathro. *Metadata: An Overview* [documento HTML]. Canberra: National Library of Australia, rev. 10 de septiembre de 1997. Disponible en: <http://www.nla.gov.au/nla/staffpaper/cathro3.html> (consultado el 8 de septiembre de 1999); Katrina Hodgson. *Metadata: Foundations, Potential and Applications* [documento HTML]. Edmonton: University of Alberta, School of Library & Information Studies, 31 de marzo de 1998, rev. 29 de abril de 1998. Disponible en: <http://www.slis.ualberta.ca/538/khodgson/metadata.htm> (consultado el 31 de octubre de 2000).

publicaciones, principalmente electrónicas, sobre metadatos <<http://www.ifla.org/II/metadata.htm>>, clasificando los enlaces que establece según los distintos formatos existentes, y también sobre bibliotecas digitales <<http://www.ifla.org/II/diglib.htm>>; o el Consorcio WWW (W3C) <<http://www.w3.org/metadata>>, donde se recogen todas las informaciones relativas a RDF y a su estado de normalización. También destacan algunos servidores institucionales a nivel nacional, como la Biblioteca Nacional de Australia que desde 1999 mantiene el sitio denominado *Meta Matters* <<http://www.nla.gov.au/meta>>, que recopila, como su propio nombre indica, asuntos relacionados con los metadatos y las bibliotecas digitales, sobre todo en el ámbito geográfico de ese país. La oficina inglesa UKOLN, compila un conjunto de páginas con enlaces a otras fuentes de información de gran interés (proyectos, foros, iniciativas, recursos, publicaciones, etc.) <<http://www.ukoln.ac.uk/metadata>>. De igual modo, podemos destacar la información emanada del grupo de trabajo sobre recuperación de información en Internet (*Resource Discovery Unit*) de la Universidad australiana de Queensland <<http://www.dstc.edu.au/RDU>> donde se da acceso a múltiples prototipos y publicaciones relacionadas con su investigación en metadatos; o el portal denominado *Biblioteca virtual sobre bibliotecas virtuais* <<http://www.prossiga.br/bibvirtual>> del CNPq (*Conselho Nacional de Desenvolvimento Científico e Tecnológico*), que ofrece recursos estructurados, tanto brasileños como internacionales, sobre bibliotecas digitales. Ameritan mención también las recopilaciones de recursos que, no con menos rigor, mantienen individualmente algunos investigadores en estos temas⁴².

⁴² Algunos de los mejores sitios en este sentido son:

- Candy Schwartz. *Metadata Resources* [documento XHTML]. Boston: Simmons College, rev. 17 de julio de 2000. Disponible en: <http://web.simmons.edu/~schwartz/mymeta.html> (consultado el 10 de febrero de 2001).
- *Metadata Standards, Crosswalks, and Standard Organizations* [documento HTML]. Charley Pennell. Queen Elizabeth II Library. Newfoundland: Memorial University of Newfoundland, rev. 7 diciembre de 2000. Disponible en: <http://www.mun.ca/library/cat/standards.htm> (consultado el 10 de febrero de 2001).

[cont.]

A tenor de la importancia de las publicaciones electrónicas sobre nuestro ámbito de investigación, es preciso comentar que el propio entramado científico de las mismas, donde la mayor parte de los trabajos que se citan —tanto en las fuentes impresas como digitales— son recursos de la Red, nos ha permitido, casi siempre acceder a los recursos originales reduciendo de este modo en nuestro trabajo, las citas de segunda mano.

En cuanto a los autores debemos destacar por sus publicaciones, tanto en formato impreso como electrónico, la labor de algunos profesionales que normalmente pertenecen a foros de investigación en este campo. Así, son especialmente significativos los trabajos de Vellucci, Baca, Weibel, Brisson, Dekkers, Miller, Guenter, Hjelm, Gill, Ianella, Oddy, Koch, Kashyap y Sheth, Denenberg, Dovey, etc., cuyas aportaciones han sido esenciales en el transcurso de la realización de este trabajo. Mención especial merecen los investigadores del UKOLN, Andrew Powell, Lorcan Dempsey y Rachel Heery cuyo influjo se ha hecho sentir principalmente, en lo relativo a la tipificación de los metadatos; así como las aportaciones de Tim Berners-Lee en torno al concepto de Web Semántica, XML y RDF, no sólo por su autoridad en el desarrollo de la WWW, sino porque la aparición

-
- Judy Ahronheim. *Judy and Magda's List of Metadata Initiatives* [documento HTML]. University of Michigan, 2 de noviembre de 1997. Disponible en: <http://www-personal.umich.edu/~jaheim/alcts/bibaccs.htm> (consultado el 12 de junio de 2001).
 - *Dave Beckett's Resource Description Framework (RDF): Resource Guide* [documento XHTML]. Bristol: University of Bristol. Institute for Learning and Research Technologies, rev. 13 de febrero de 2001. Disponible en: <http://www.ilrt.bris.ac.uk/discovery/rdf/resources> (consultado el 13 de febrero de 2001).

Finalmente, nos gustaría revelar que a raíz de esta tesis doctoral, y ante la carencia de un sitio en español semejante, que comente y aglutine los recursos y la investigación sobre este tema, el 3 de junio de 2001 registramos el dominio: <<http://www.metadatos.net>> donde esperamos mantener un recurso de información y referencia, en español, sobre metadatos y bibliotecas digitales. Además de esta iniciativa personal, dentro del ámbito de SEDIC y del grupo de trabajo sobre Normalización <http://www.sedic.es/grupo_internet.htm> creado en mayo de 2001, se está traduciendo al español el sitio de la DCMI que se publicará a principios del año 2002, en el URL <http://www.dublincore.sedic.es>, y que aspira a ser un *mirror* de este sitio, de forma semejante a los que existen en Australia y en el Reino Unido.

de su libro *Tejiendo la Red*⁴³ en el año 2000, corroboró muchas de nuestras hipótesis y aportó seguridad a nuestra tesis doctoral. En cuanto al capítulo dedicado a la Biblioteca Virtual *Miguel de Cervantes* han sido decisivas las publicaciones de Alex Bia para poder adaptar un planteamiento de metadatos a la realidad tecnológica de dicha biblioteca.

Hemos advertido dos tendencias distintas entre los autores que hemos tenido en cuenta: por un lado aquellos sectores más tradicionales aferrados al formato MARC y al mundo catalográfico impreso, que piensan que Internet no cambiará substancialmente su función de hacer accesible el conocimiento y ven en la Red una manera de perpetuar el trabajo metódico bibliotecario partiendo de que la organización de la Red es *caótica* (Kuny, Cleveland, etc. e incluso desde un punto de vista más técnico, Olson o Jul). Por otro lado, un sector, extremadamente optimista, que tiene la visión onírica de Internet como la *gran biblioteca digital* (v. gr. Ioanid y Bowman o Lesk). El influjo de ambas visiones se puede identificar en distintas partes de esta investigación; sin embargo, hemos tratado de mediar entre ambas posturas para presentar un panorama más realista y concreto de la función del bibliotecario frente Internet, así como de los riesgos y oportunidades que comportan los metadatos y la estructuración de la información en las bibliotecas digitales.

Respecto a la bibliografía, sólo nos resta manifestar la inminente transformación de la información relativa al tema de los metadatos, teniendo en cuenta el vertiginoso desarrollo de los estándares de Internet en los últimos años. Por ello, hemos intentado acceder a la información más reciente sin olvidar los documentos fundamentales en la definición y conceptualización de los metadatos así como en su desarrollo o evolución "histórica". Por esta misma razón, hemos accedido a los recursos electrónicos varias veces en el transcurso de esta investigación

⁴³ Tim Berners Lee. *Tejiendo la red: el inventor del World Wide Web nos descubre su origen*. Madrid: Siglo veintiuno, 2000. Es la versión española de la obra original publicada en septiembre de 1999: Tim Berners-Lee with Mark Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web, by its Inventor*. San Francisco [New York]: Harper, 1999.

consignando, en la medida de lo posible, la fecha de acceso (en las notas al pie) y la fecha de comprobación final (en la nota introductoria a la bibliografía).

Por último, debemos señalar que los metadatos gozan todavía de esa situación de excepción de todo lo que empieza desde cero; empero no podemos olvidar los problemas a los que se han enfrentado los bibliotecarios y documentalistas en los últimos 30 años. Así, se ha hecho inevitable la utilización de obras clásicas sobre recuperación de información (Rijsbergen, Blair, etc.), cuya consulta y referencia son obligadas en un trabajo de estas características.

PARTE I:

De la información a la metainformación: metadatos

3. Metadatos: concepto, fundamento, aplicación y tipología. 3.1. Precisiones terminológicas y uso histórico del término. 3.2. Concepto. 3.3. Fundamento de la metainformación: el Documento como Objeto. 3.4. Aplicación de los metadatos. 3.5. Tipos de metadatos. **4. El problema de la asignación de metadatos.** 4.1. De la catalogación a la asignación de metadatos: reflexiones en torno al proceso técnico de los recursos electrónicos. 4.2. Quién crea los metadatos. 4.3. Cómo asociar metadatos a los recursos de información electrónica: herramientas y aplicaciones.

CAPÍTULO 3:

METADATOS: CONCEPTO, FUNDAMENTO, APLICACIÓN Y TIPOLOGÍA

En este capítulo nos aproximaremos al estado de la investigación sobre el tema de los metadatos, abordando para ello cuestiones teóricas sobre el concepto, la justificación y fundamento del mismo, su caracterización, uso y tipología. De esta forma pretendemos construir aquí un acervo de reflexiones iniciales que nos encaminen a demostrar la hipótesis —a partir del concepto, en este caso— de la necesidad de contar con metadatos para optimizar la búsqueda y recuperación de información en Internet y más concretamente, en el contexto de las bibliotecas digitales.

3.1. Precisiones terminológicas y uso histórico del término

Como término y concepto principal de este trabajo de investigación, pretendemos utilizar "metadatos" como una palabra descriptiva *per se*. Sin embargo es preciso comentar que su excesiva utilización, sus diferentes significados, dependiendo del contexto o del colectivo que la utilice, y el hecho de ser una palabra *de moda* en el ámbito de la Documentación, hacen que tenga un significado bastante vago si no se le añade un calificador que precise su alcance, contenido o campo de aplicación.

Podríamos hacer una reflexión etimológica que nos lleve a enjuiciar la carga semántica de este término. De manera sistemática cabe pensar que, a pesar de su procedencia anglosajona (*metadata*), la palabra metadatos está compuesta por el

prefijo griego **meta-** [μετα-], que según el Diccionario de la Real Academia Española¹ (DRAE), significa: *junto a, después de, entre, con*, más el término de origen latino **dato/s** [datum], cuyo significado no presenta ninguna ambigüedad: *representación de una información de manera adecuada para su tratamiento por un ordenador*², aunque en el contexto de la Documentación automatizada tiene un sentido muy amplio:

*Dato/s —data, en inglés— son distintas piezas de información formateadas de una manera especial. Todos los software se dividen en dos categorías generales: datos y programas. Los programas son colecciones de instrucciones para manipular los datos. Los datos pueden existir en una variedad de formas —como números o texto en papel, como bits y bytes almacenados en una memoria electrónica o como hechos acopiados en la memoria de una persona*³[....]

Según todo esto, estricta y etimológicamente los metadatos serían:

datos {junto a | después de | entre | con...} los datos.

Por otra parte, si partimos del concepto de metalenguaje entendido, también según el DRAE, como *lenguaje cuando se usa para hablar del lenguaje mismo*, de forma paralela, podríamos decir que los metadatos son datos cuando se usan para hablar de ellos mismos, y de ahí que, de manera generalizada y válida en cualquier contexto, se defina los metadatos como "datos sobre los datos" o "datos acerca de

¹ *Diccionario de la lengua española: edición electrónica. Versión 21.1.0* [CD-ROM]. Madrid: Espasa-Calpe, 1995. En la versión impresa de esta obra de referencia: *Diccionario de la lengua española*. 21ª ed. Madrid: Real Academia Española, 1992, Tomo II, p. 1363.

² *Ibid.* Tomo I, p. 663 (acepción nº 3 de la voz *dato*).

³ Esta es la definición que da, al respecto de la palabra "datos" (*data*), Webopedia, una fuente electrónica de referencia especializada en conceptos informáticos, sobre todo, relacionados con la WWW. *Webopedia: Definition and Links* [documento HTML]. INT Media Group Incorporated, 2000. Disponible en: <http://webopedia.internet.com/TERM/d/data.html> (consultado el 6 de noviembre de 2000).

datos" y metainformación como "información sobre la información"⁴. A esta conclusión también podemos llegar si tenemos en cuenta el significado que da Berners Lee a "meta-": *prefijo que indica que algo se aplica a sí mismo; por ejemplo, una metareunión es una reunión acerca de reuniones*⁵. Teniendo en cuenta estas reflexiones, no cabe ninguna duda que los metadatos son también datos, pero datos representacionales, que añadidos a la propia información adquieren el valor semántico para sustituirla o representarla.

A pesar de la asociación o paralelismo al término metalenguaje, y de la definición de Berners-Lee, el prefijo "meta-" tiene un significado establecido y diferente cuando se usa como afijo antepuesto a otras palabras distintas a datos. Por ello, el significado de metadatos no puede inferirse estrictamente el sentido que adoptan otros términos al anteponerles "meta". Según esto, podríamos también desechar algunos significados para la voz metadatos: por ejemplo, no son datos que hayan cambiado de posición o de forma, que estén alterados o transpuestos, como se podría colegir del significado de metamorfosis; ni son datos abstractos, sutiles u oscuros, más allá de lo físico, como implica el prefijo meta- en metafísica; tampoco son datos detrás de algo o en la parte trasera de algo, como se podría derivar de términos como metatórax.

Es preciso añadir además, que en la actualidad existe en Internet un uso y abuso de la partícula *meta-* antepuesta a términos de todo tipo. Así encontramos, por ejemplo: metainvestigación, metagrupo, metaguía, metabase, metaíndice o

⁴ Higgins da un significado más profundo al término metainformación: un entendimiento de alto nivel que se tiene sobre un ítem de información, como la credibilidad de las fuentes para la toma de decisiones. En cierto sentido, como veremos, este significado puede inferirse también para los metadatos. Margaret Higgins. *Meta-Information and Time: Factors in Human Decision Making. Journal of the American Society for Information Science*, 1999, vol. 50, nº 2, p. 132-133.

⁵ Tim Berners Lee. *Tejiendo... Op. cit.*, p. 225.

metadirectorio, metalista, e incluso metaenciclopedia, entre otros⁶. De todas estos "meta-x" los conceptos más arraigados y que no tienen lugar a dudas, independientemente de la carga etimológica del prefijo, son principalmente dos:

- El primero de ellos, *metabuscador*, no tiene estrictamente ninguna relación con los metadatos, este término se utiliza simplemente para aludir a aquellos sistemas de búsqueda que lanzan las preguntas a distintas bases de datos, de diferentes servicios de recuperación de información en la Red, resumiendo los resultados y evitando duplicados, como es el caso del propio Metacrawler <<http://www.metacrawler.com>>.
- El segundo de los términos más utilizados en el contexto de la información Web es *metaetiqueta (metatag)*, y en este caso, sí tiene una gran relación con los metadatos. En esencia, las metaetiquetas provienen del propio HTML en que están constituidos, hasta ahora, la mayor parte de los documentos Web. Son las etiquetas que sirven para determinar, en una cabecera de un documento, el elemento META. En la especificación en vigor de HTML en el momento de realización de esta investigación, al hablar de la etiqueta <META>, se mencionan explícitamente los metadatos como *información*

⁶ Algunos URLs que demuestran la utilización de estos nuevos términos en la Red son, por ejemplo:

- Metaresearch <<http://www.metaresearch.org>>; Metagroup <<http://www.metagroup.com>> ;
- Meta-guide <<http://www.meta.fgov.be>>; Metaíndice <<http://gsulaw.gsu.edu/metaindex>>;
- Metabase <<http://www.metabase.net>> que es una "base de bases de datos" sobre medio ambiente y desarrollo sostenible en Centroamérica;
- Metalista <<http://www.meta-list.org>> No es más que un servicio de listas de distribución sobre ciencia y religión que ha tomado el prefijo "Meta-" como nombre para denominar dicho servicio incluyendo además en su presentación el significado del mismo. Un caso semejante lo encontramos en la empresa MetaHTML <<http://www.metahtml.com>> que ha adoptado dicho prefijo para su nombre comercial.
- Meta-enciclopedia <<http://www.ditext.com/encyc/frame.html>> Se denomina así a este servicio concreto que permite hacer búsquedas en varias enciclopedias de filosofía a la vez, emulando el funcionamiento de los propios metabuscadores.

sobre el documento en lugar del contenido del documento⁷. Sin embargo las metaetiquetas son la materialización, en el marcado HTML, del concepto de metadatos, y su explicación y uso en la WWW, tiene siempre un tono de divulgación en torno a cómo hacer más visibles las páginas Web ante los motores de búsqueda⁸.

La definición de los metadatos, que etimológicamente parecía tan sencilla, encierra una gran complejidad. A pesar de que el término parece una mera aplicación de la etimología clásica, para denominar la necesidad de estructurar y describir el contenido de los recursos de información de la Red, su origen y su utilización histórica, distan mucho de esta afirmación, como veremos a continuación.

Pese a que se ha difundido a raíz de Internet y como soporte al acceso a la información en la World-Wide Web, el término metadatos es anterior al Web, y fue supuestamente acuñado, en los años 60 por Jack E. Myers para describir eficazmente conjuntos de datos⁹. En 1969, después de definir la arquitectura de lo que

⁷ "...Information about a document rather than document content". Esta definición de metadatos aparece explícitamente en el capítulo 7 de la especificación del HTML 4.01, la última en vigor, en el momento de realización de esta investigación. The Global Structure of an HTML Document. En: World Wide Web Consortium. *HTML 4.01 Specification W3C Recommendation 24 December 1999* [documento HTML]. Dave Raggets, Arnaud Le Hors, Ian Jacobs, eds. WWW Consortium, 24 December 1999. Disponible en: <http://www.w3.org/TR/html4/struct/global.html#h-7.4.4> (consultado el 19 de octubre de 2000).

⁸ Sobre el concepto de metaetiqueta incidiremos en otros capítulos de este trabajo. Vid. 4.3.1. y 5.2.1.1. Para una definición y diferenciación exhaustiva con respecto a los metadatos, Vid. Glosario, *metatag*.

⁹ Denis Howe. *FOLDOC: Free Online Dictionary of Computing* [documento HTML], rev. 29 de junio de 1998. Disponible en: <http://wombat.doc.ic.ac.uk> (consultado el 6 de julio de 1998), también disponible en: <http://www.de.easynet.net/resources/foldoc/foldoc.cgi?Metadata> (consultado el 31 de octubre de 1999). En este mismo diccionario, se distinguen dos voces: "Metadata", y "meta-data" argumentando que cuando Myers acuñó la palabra "metadata" no descubrió ningún uso especial de dicho término, sino que lo utilizó para designar su empresa, como hemos explicado. Sin embargo, se define meta-data de la siguiente forma: *Datos sobre los datos. En el procesamiento de datos, los metadatos son datos que proveen información o documentación sobre otros datos gestionados dentro de una aplicación o contexto determinado. Por ejemplo, los metadatos aportarán información sobre* [cont.]

posteriormente denominaría "*metamodel*", Myers utilizó por primera vez el término *metadata*, aunque tal palabra no aparecería impresa hasta 1973 en un folleto relativo a su producto. En un principio utilizó el término con la intención de que no significase nada en particular, sino que fuese un término fácil de recordar y en definitiva, comercial, un término nuevo, que utilizaría para designar su empresa: *Metadata Information Partners*.

La marca *METADATA*® se registró en 1986 en la Oficina de Patentes y Marcas de los Estados Unidos con el registro N°1.409.206 y es una marca comercial que pertenece a la *Metadata Company*¹⁰. La propiedad de la marca quedó fuera de duda en 1991, cuando se utilizaba como nombre comercial, siempre como nombre propio con mayúscula inicial, evocando a la *Metadata Company*. Dicha empresa pretendía que, si el objetivo de utilizar el término "metadata" era hablar de "datos sobre los datos", se debía utilizar: "*meta data*" (meta datos) o "*meta-data*" (meta-datos). Sin embargo el desarrollo de la teoría de metadatos, tal y como la reflejamos en este trabajo, así como el amplio despliegue de modelos y estándares en los últimos años, demostrará un uso más convencional del término, vinculado a la descripción de recursos en Internet y totalmente independiente de dicha marca comercial. De forma semejante, otras empresas del sector han adoptado también el nombre, tal es el caso, por ejemplo, de una franquicia de la empresa Microsiga Software en Uberlandia, Minas Gerais (Brasil) que desde julio de 1999 se denomina Metadata <<http://www.metadata.com.br>> y se dedica a la consultoría de soluciones para el negocio electrónico (*e-business*).

elementos de datos o atributos (nombre, tamaño, tipo de datos, etc.) e información sobre los registros o la estructura de datos (extensión, cómo están asociados, a quién pertenecen, etc.). Los metadatos pueden incluir información descriptiva sobre el contexto, calidad y condiciones o características de los datos.

¹⁰ Metadata vs. Meta-data/Meta Data. [documento HTML]. En: *Metadata Company: Innovators in Data Management Technology*. Metadata Company, rev. 6 de marzo de 2000. Disponible en: <http://www.meta-data.com/word.htm> (consultado el 31 de octubre de 2000).

Este término comienza a aparecer, de forma frecuente, en los años 80 en la literatura sobre Sistemas de Gestión de Bases de Datos (SGBD/DBMS). De hecho, la primera publicación que encontramos con el término *metadata* en el título en la base de datos LISA, es del año 82, y define los metadatos, en el contexto de los SGBD, de la siguiente manera:

Los metadatos son información sobre datos numéricos [...] un sistema de metadatos funciona como un equivalente paralelo a los sistemas de datos numéricos de tal forma que los analistas, los que toman decisiones, los que resuelven problemas y los gestores de sistemas, aprendan suficiente sobre los datos numéricos para resaltar la probabilidad de su uso válido y apropiado¹¹.

En aquella época, la voz metadatos se usaba para describir la información que documentaba las características de la información contenida en las bases de datos. En el dominio de los SGBD, los ficheros de datos son los que almacenan la información en sí misma, mientras que otros archivos, tales como el índice y los diccionarios de datos, almacenan información administrativa conocida como metadatos. El ordenador era el escenario tanto para la información que se debía describir, como para los datos descriptivos, y por lo tanto los metadatos funcionaban en un contexto totalmente electrónico. De ahí pues, quizás, su origen de vincularlos a un mundo de información estrictamente electrónica, numérica o digital.

Mientras tanto, en los mismos 80s, en el resto de la literatura profesional bibliotecaria se utilizaban expresiones como: descripción bibliográfica, datos catalográficos, datos bibliográficos, o simplemente catalogación, para denominar a esos datos sobre los datos cuando, tanto el objeto catalogado (monografía impresa, mapa, videgrabación, música impresa, etc.), como el registro que lo describía, estaban en formato no electrónico. Inclusive, se seguía denominando así a la

¹¹ M. David Liston, L. James Dolby. Metadata Systems for Integrated Access to Numeric Data Files. *Drexel Library Quarterly*, Summer Fall 1982, vol. 18, nº 3-4, p. 147-160. Definición extraída del resumen de este artículo en la base de datos LISA.

información referencial o descriptiva, cuando el registro bibliográfico se realizaba de forma electrónica en formato MARC, aún cuando el material descrito era un fichero de ordenador local (disquete o CD-ROM). Sin embargo, cuando los catalogadores empiezan a describir recursos electrónicos de la Red, el formato MARC comienza a entenderse o proponerse como un modelo más de metainformación.

Aunque es difícil datar este cambio en la concepción de la descripción de recursos de información en el contexto documental, y la forma de designar dicho proceso descriptivo, podríamos apuntar un hito significativo en la tradición normativa bibliotecaria: el momento en que las propias normas ISBD para archivos de ordenador —ISBD (CF)— se revisan en 1997, cambiando, de manera especialmente significativa, la denominación del tipo de material a "recursos electrónicos" (*Electronic Resources: ER*) y ampliando la tipología de ficheros informáticos a los archivos electrónicos remotos de Internet¹². Las ISBD (ER) no mencionan en ningún momento el concepto de metadatos, ni siquiera como fuente para la catalogación. Por ello y desde nuestro punto de vista, la utilización de este término, más que deberse a dicho cambio en la aprehensión de las normas ISBD aplicadas al material electrónico, tiene su esencia en la convergencia de la Biblioteconomía y Documentación, con el mundo mucho más amplio y ambiguo de las Tecnologías de la Información y de las Comunicaciones (TIC) vinculadas al desarrollo de Internet, donde concurren otros actores como la informática, las políticas de información y la organización del conocimiento en general. Todas estas circunstancias han servido para contemporizar la práctica de la catalogación tradicional con el nuevo concepto, de tal forma que "metadatos" se ha convertido en

¹² IFLA. *ISBD (ER) International Standard Bibliographic Description for Electronic Resources: Revised from the ISBD (CF) International Standard Bibliographic Description for Computer Files*. München: Saur, 1997. También disponible en Internet en: <http://www.ifla.org/VII/s13/pubs/isbd.htm> (consultado el 25 de abril de 2001). Pedraza analiza la norma ISBD (ER) prestando especial atención a la evolución de la normativa para la descripción de recursos electrónicos desde la aparición de las ISBD. Manuel José Pedraza García. Las ISBD (ER): Adecuación normativa a un panorama cambiante. *Scire*, enero-junio 1998, vol. 4, nº 1, p. 45-55.

un término aceptado en todas las disciplinas vinculadas, de una u otra forma, a la organización de la información y/o el conocimiento en Internet.

En cualquier caso, hasta mediados de los 90s (en 1995, se celebra el primer *workshop* del Dublin Core), encontramos el término "metadatos", en comunidades implicadas en la gestión e interoperabilidad de datos geoespaciales, y en general en el diseño y mantenimiento de sistemas de gestión de datos. Para estas primeras comunidades de gestores de información que utilizaban metainformación de forma semejante a como los entendemos ahora, los metadatos se referían a un conjunto de normas industriales, junto a una documentación adicional interna y externa y otros datos necesarios para la identificación, representación, interoperabilidad, gestión técnica, funcionamiento y uso de los datos albergados en un sistema de información. Ese componente normativo de estructuración y definición de la información es el que persiste en el significado que se ha conferido a principios del siglo XXI, y que damos aquí, a los metadatos. Esta reflexión histórica sobre la evolución del término, así como su utilización en el ámbito bibliotecario y de la catalogación, será fundamental para la explicación que realizaremos a continuación en torno a las corrientes conceptuales sobre los metadatos, que hemos podido advertir a lo largo de esta tesis.

3.2. Concepto

Son muchos los autores que califican negativamente al término (metadatos) o a su definición (datos sobre los datos) y se refieren, tanto al término como al concepto, con distintos calificativos que evidencian su ambigüedad o "torpeza" designativa. Sobre el vocablo "metadatos", Dempsey¹³ dice que es un término *torpe*, que sirve

¹³ *Metadata, a clumsy term, is data which helps in the identification, location or description of a source*). Lorcan Dempsey. Meta Detectors [documento HTML]. *Ariadne*, 20-May-1996, issue 3. Disponible en: <http://www.ariadne.ac.uk/issue3/metadata> (consultado el 25 de junio de 1998). Cfr. *infr.* Caplan, nota 32.

para designar los datos que ayudarán en la identificación, localización o descripción de un recurso. Gorman¹⁴ es aún más severo al decir que *metadatos es un término selecto o caprichoso para una forma de catalogación inferior*. La definición más arraigada asociada al término *metadatos*, esto es, "datos sobre los datos", tampoco es muy afortunada para delimitar el concepto. A pesar de ser una definición exacta, es simplista o falta de detalle, sibilina para Ducloy¹⁵, y *añade muy poco a la comprensión del concepto*¹⁶.

Además de la definición trivial y confusa, derivada de la etimología o de la semejanza morfológica con otros términos, *datos sobre los datos (cosas sobre cosas, para Dovey¹⁷)*, es preciso matizar bien el significado de los metadatos y sus características, antes de abordar ninguna reflexión teórica sobre ellos o sobre su adecuación como herramienta descriptiva para mejorar la localización y la

¹⁴ *A fancy name for an inferior form of cataloguing*. Michael Gorman. From Card Catalogues to WebPACS [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 1 de noviembre de 2000, rev. 16 de noviembre de 2000. Disponible en: http://www.unicamp.br/bc/gorman_paper.html (consultado el 21 de febrero de 2001). *Cfr. infr.* Nota 32.

¹⁵ Estamos de acuerdo con Ducloy al calificar la definición de *datos sobre los datos* como *muy sibilina*. Jacques Ducloy, Catherine Lupovici, Elizabeth Cherhal. Les metadonnées et le catalogage des documents numériques [documento HTML]. En: *Minutes des rencontres BIBLIO-FR (1998. Caen)*. Caen: Université de Caen, Laboratoire GREYC, abril de 1998, rev. 15 de enero de 1999. Disponible en: <http://www.info.unicaen.fr/bnum/biblio-fr/rencontres98/minutes/metadonnees/texte.html> (consultado el 22 de febrero de 2001). Desde nuestro punto de vista, el hecho de partir de una definición tan hermética y confusa, a pesar de su precisión, hace que se asimilen los metadatos a la catalogación lo que, como veremos, no es del todo exacto (*Cfr.* 4.1).

¹⁶ *While accurate, this definition is so simplistic that its adds very little to our understanding of the concept*. Sherry L. Vellucci. Metadata and Authority Control. *Library Resources & Technical Service*, 2000, vol. 44, nº 1, p. 33.

¹⁷ Matthew J. Dovey "Stuff" about "Stuff"- the Differing Meanings of "Metadata". *Vine (Theme issue: Metadata. Part 1)*, 1999, nº 116, p. 6-13.

recuperación de información en Internet, ni sobre su utilización como base estructural de las emergentes bibliotecas digitales.

Partiendo de esa definición lacónica y manida en la literatura sobre este tema, *datos sobre datos o información acerca de la información*, cabe pensar que es un concepto sencillo y bien delimitado. Sin embargo, y como demostraremos a continuación, no existe un consenso conceptual sobre este término. Ni siquiera es posible colegir una aproximación a los metadatos que pueda cubrir todos los objetos de información y que sea válida para todos los contextos en que podemos hablar, de una u otra forma, de ellos. Con este planteamiento, uno de los primeros retos que tiene que asumir la investigación sobre metadatos —y por ello lo hacemos aquí— es el establecimiento de una definición inequívoca y ajustada a cada entorno, que tenga sin embargo un carácter universal, más allá de la mera organización bibliotecaria.

La primera dificultad que encontramos para establecer una definición unívoca es la cobertura diferente del término según el colectivo que lo utilice. Los informáticos usan la palabra metadatos para referirse a una cosa y los bibliotecarios lo hacen para referirse a otra, además de otros sectores, al margen de estos dos mundos, que también la utilizan. Teniendo en cuenta las opiniones del IEEE (*Institute of Electrical and Electronics Engineers*) —que cita Goodchild en uno de los primeros talleres sobre metadatos¹⁸—, para un ingeniero de sistemas, metadatos significa un nivel físico de información como los nombres de los ficheros o formatos, los tipos de datos, etc., es decir, lo que se necesita para decodificar una secuencia de *bytes* en elementos reconocidos por un lenguaje de programación de propósito general. Para un gestor de bases de datos, sin embargo, los metadatos pueden ser el contenido de un esquema, o sea, los nombres para todas las clases de objetos de datos

¹⁸ Michael F. Goodchild. *Alexandria Digital Library: Report on Workshop on Metadata* [documento HTML]. Santa Bárbara: University of California, 8 de noviembre de 1995. Disponible en: http://www.alexandria.ucsb.edu/public-documents/metadata/metadata_ws.html (consultado el 18 de junio de 1998).

dentro de la base, una sentencia concreta de todos sus atributos y de todas las relaciones entre ellos, así como una caracterización de las preguntas que pueden hacerse a una base de datos determinada. Para un físico, por ejemplo, los metadatos pueden ser un número que se usará en la transformación de los datos, o pueden significar una descripción en lenguaje natural de los procesos de medida. Por otra parte, para un neófito que está estudiando un nuevo dominio de información, pueden suponer simplemente una guía para encontrar más información. Para un bibliotecario-documentalista, ámbito profesional en el que se inserta este trabajo, podemos enunciar que los metadatos son un tipo de datos que usamos para describir el contenido, estructura, representación y contexto de algún conjunto de datos específico.

En la misma línea de diferenciación de contextos de utilización de este concepto, Gilliland-Swetland menciona algunos ejemplos de situaciones más específicas donde puede adoptarse el vocablo metadatos. Esta autora hace especial hincapié en la amplitud de significado cuando se utiliza en contextos *menos tradicionales*¹⁹. Cita, por ejemplo, el caso de un proveedor de recursos de Internet que podría utilizar el término metadatos refiriéndose a la información codificada como metaetiquetas HTML de tal forma que el sitio sea más fácil de encontrar; o un digitalizador de imágenes para quien los metadatos pueden ser la información que se introduce en el encabezamiento de una ficha digital que registra información sobre la imagen; o el caso de un archivero de datos de ciencias sociales que podría usar el término para referirse a los sistemas y a la documentación necesaria para interpretar una cinta magnética de datos no procesados; o aquellos profesionales que se encargan de archivar datos electrónicos, y que pueden utilizar metadatos para designar la

¹⁹ Anne Gilliland entiende por *ámbitos más tradicionales* aquellos relativos a los especialistas de información sobre el patrimonio cultural, tales como bibliotecarios, archiveros y administradores de museos. Anne J. Gilliland-Swetland. La definición de los metadatos. En: *Introducción a los metadatos: vías a la información digital*. Murtha Baca, ed. Los Angeles: J. Paul Getty Trust, 1999, p. 1-2.

información contextual, de procesamiento y de uso necesaria para identificar y documentar la aplicación, la autenticidad y la integridad de un registro en un sistema electrónico de almacenamiento de información. Todas estas interpretaciones diversas sirven para ilustrar la ambigüedad o amplitud del término según el ámbito profesional y/o científico en que se utilice.

Otro de los aspectos importantes en la definición de los metadatos, es el nivel de aplicación de los mismos, es decir, si su implantación es local, en un dominio o comunidad concretos, o por el contrario, si se utilizan en un contexto amplio de interoperabilidad en la Red. En este sentido, los metadatos, se aplican²⁰, se han aplicado y se aplicarán a muchos dominios diferentes, es decir, a diversas áreas que requerirán enfoques diferentes, como por ejemplo²¹:

- Recursos de información local, estatal o federal, tales como publicaciones gubernamentales, o el sitio Web de una agencia federal.
- Información museística: exposiciones de patrimonio cultural compuestas de un catálogo de exposición, imágenes, textos murales y arte folklórico.
- Libros, revistas, periódicos, o colecciones de materiales relacionados pertenecientes a bibliotecas digitales. Este será el entorno informacional que

²⁰ La aplicación de los metadatos se tratará más profusamente en el apartado 3.3. *Vid. infr.*

²¹ Hemos tomado como referencia (aunque sin respetar el orden) la casuística de contextos que recogía la empresa *Blue Angel Technologies* en 1998, cuando se estaba gestando el concepto, variedad y posibilidad de los metadatos, ya que sus perspectivas de mercado se basaban en la investigación de los diversos estándares de metainformación. *Blue Angel Technologies. Metadata Explained* [documento HTML]. *News and Info*, 1996-1998, rev. 10 de abril de 1998. Disponible en: <http://www.bluangel.com/NewsAndInfo/faqs/metadata.htm> (consultado el 19 de junio de 1998). Actualmente (febrero 2001) la empresa *Blue Angel* <<http://www.BlueAngeltech.com>>, comercializa —quizás también fruto de sus estudios de mercado en el ámbito de los metadatos— el software *MetaStar*, cuyo reclamo comercial se funda en el hecho de ser una herramienta basada en XML y vinculada al modelo B2B (*Business to Business*) del comercio electrónico, y a su potencial como herramienta para compartir información en bibliotecas digitales <<http://www.BlueAngeltech.com/products/Solutions/lib.htm>>. *Vid.* 4.3.2.3.

más nos interesa para este trabajo, ya que como venimos diciendo desde la introducción, la edición electrónica y "las" bibliotecas digitales, son una realidad a la que se tiene que enfrentar el profesional de la información con el esfuerzo normalizador y de estructuración que ha caracterizado su trabajo durante siglos, y que ahora debe adecuar a este nuevo contexto informativo.

Además de estas áreas de información, que señalaba la empresa *Blue Angel* y que han desarrollado una fuerte y ya consolidada trayectoria en el ámbito de los metadatos —como se verá en otras partes de este trabajo—, también es importante señalar otros colectivos que trabajan de manera semejante, definiendo modelos estructurales aplicables a sus tipos de información o desarrollando formatos de metainformación *ad hoc*. Como por ejemplo el de los sistemas de información geográfica y geoespacial, o el mundo de la información de archivo, el comercio electrónico, el colectivo de la información vinculada a la educación, así como otros menos comunes o más específicos, como la aplicación a la información oceanográfica²².

En realidad, casi todas las comunidades científicas que, de una u otra forma, desarrollan sistemas de información basados en una arquitectura de Red (Internet/Intranet) se han planteado, explícita o tácitamente, una estructuración de su información electrónica en forma de metadatos. Por su parte, los sistemas de información tradicional, como los catálogos de bibliotecas, aunque no han denominado a sus representaciones de la información o registros bibliográficos "metadatos" han adoptado sistemas de descripción de los recursos con una finalidad idéntica a la de éstos: crear estructuras para organizar la información. En cualquier caso, se ha convertido en un término sobrecargado, que se utiliza en distintos

²² Tal es el caso por ejemplo del trabajo desarrollado por Ivette Patricia Jiménez Guardiola y Carlos Arturo Parra Llanos en la Universidad Tecnológica de Bolívar de Cartagena de Indias. *El uso de RDF en el diseño de sistemas de información en la Web y su aplicación en el prototipo Oceandata*. 2000, 43 p., (ejemplar cedido por Carlos Parra, coaut.).

contextos, por comunidades de áreas científicas distintas, con unos objetivos y una motivación diferente. Sin embargo, todas estas áreas comparten, en la mayoría de los casos, herramientas, estándares y problemas en torno a los metadatos.

Dempsey y Heery²³ hacen una aproximación genérica al concepto de metadatos a través de un discurso definitorio basado en ejemplos, que puede coadyuvar a resolver el aspecto del dominio y/o nivel de aplicación, y que constituye asimismo, una primera aproximación intuitiva al concepto que estamos tratando. En esta visión, a través de contextos prototipo donde podemos encontrar datos sobre los datos, los autores explican qué es lo que podríamos entender por metadatos en diferentes ámbitos informativos, es decir, qué pueden parecer los metadatos, más que lo que efectivamente son, teniendo en cuenta el colectivo que los utiliza y el uso que se hace de ellos.

En esta línea, podemos pensar que casi todos, por no decir todos, los recursos de información incluyen en sí mismos datos que los describen. La mayoría de los documentos tradicionales impresos, conllevan datos autodescriptivos, *v. gr.* título, autor, incluso un resumen del contenido (pensemos por ejemplo en los artículos de publicaciones científicas). De igual forma, en las versiones digitales de los documentos, esos datos descriptivos de la propia publicación, se pueden señalar a través de etiquetas específicas (por ejemplo, <TITLE> en HTML), que permiten incluir los metadatos en la parte del encabezamiento del documento Web (<HEAD>); igualmente, los ficheros de imagen PNG (*Portable Network Graphics*) o TIFF (*Tagged Image File Format*) pueden admitir también, alguna descripción estructurada de los datos de la propia imagen.

Un tipo particular de información, dentro de la casuística que señalan estos mismos autores, son los mensajes de correo electrónico en Internet. Los *emails*

²³ Lorcan Dempsey and Rachel Heery. Metadata: A Current View of Practice and Issues. *Journal of Documentation*, March 1998, vol. 54, n° 2, p. 146-147.

siempre contienen, en la cabecera del mensaje, metainformación simple de tipo atributo-valor (fundamentalmente datos relativos al autor, destinatario, asunto y fecha), que pueden utilizarse para buscar por nombre del autor, o por materia, por ejemplo, en sistemas de acceso a mensajes de listas de distribución. Este tipo de comunicación de mensajería electrónica podría exceder la tipología documental electrónica susceptible de aplicar metadatos, sin embargo, si pensamos en sistemas de información integrales en contextos Intranet, la comunicación electrónica constituye una información muy valiosa en términos de conocimiento tácito distribuido en una organización, insustituible, por ejemplo, para la vigilancia tecnológica.

El programa de Bibliotecas electrónicas (eLib²⁴) ha establecido varias pasarelas temáticas (*subject gateways*) que constituyen verdaderas bibliotecas digitales o servicios de información de calidad, gracias a la selección y organización de sus recursos según áreas temáticas. Los registros que conforman las bases de datos de eLib tienen información descriptiva suficiente (metadatos), similar a la de las bases de datos tradicionales, es decir, con indización y resumen, además de algunos atributos adicionales sobre características técnicas.

Los servidores Z39.50²⁵ también pueden proporcionar datos técnicos sobre los tipos de búsqueda que soportan, sobre las bases de datos a las que posibilitan el

²⁴ Vid. Joint Information Systems Committee (JISC). *eLib: Electronic Library Program* [documento HTML]. Bath: University, UKOLN, rev. 2 de julio de 1998. Disponible en: <http://www.ukoln.ac.uk/services/elib> (consultado el 17 de julio de 1998). Algunos de estos *subject-based services*, o *subject gateways* del proyecto eLib son: ADAM, EEVL, IHR-Info, OMNI, SOSIG, etc. todos ellos desarrollados en el contexto de la educación superior británica, de los que tendremos oportunidad de hablar en otras partes de esta tesis doctoral, fundamentalmente en el capítulo dedicado a las bibliotecas digitales. Vid. 8.2.2. y Glosario.

²⁵ Z39.50 es un protocolo estándar para la recuperación de información, específicamente diseñado para soportar la recuperación de servidores de información distribuida. Tendremos oportunidad de volver a hablar de este estándar ANSI/NISO, a lo largo de este trabajo, así como de su integración y/o evolución a raíz de los metadatos y de metalenguajes de estructuración de información como XML.

acceso, o sobre las propias condiciones de acceso, esto es, pueden incluir metadatos técnicos²⁶.

Otro contexto en el que podemos hablar de metadatos es el de aquellos organismos interesados en poseer registros más que en obtener datos, el caso de las instituciones que se encargan de tener *testimonios escritos sobre hechos del pasado que puedan considerarse de forma fidedigna en el futuro*²⁷. Nos estamos refiriendo sin duda al mundo de los archivos, oficinas de registro, organizaciones gubernamentales y empresas, donde cada ítem de información requiere metadatos descriptivos, de tal forma que el usuario pueda identificar que los registros están completos, identificables y autorizados; requieren asimismo, datos sobre el origen, uso histórico, propiedad, condiciones de utilización establecidas por el autor; características de conservación o preservación, etc.

En general, cualquier agrupación de datos que pretenda tener una utilidad, como por ejemplo un conjunto de datos estadísticos²⁸, precisa incluir información descriptiva sobre los datos, para que estos sean efectivamente útiles. Esto ocurre también en los Sistemas de Información Geográfica (SIG), que necesitan aportar información (metadatos) sobre las aplicaciones de los datos en cada situación. De

²⁶ *Vid. infr.* Tipos de metadatos 3.5.

²⁷ Lorcan Dempsey and Rachel Heery. Metadata: A Current View... *Op. cit.*, p. 147.

²⁸ Por ejemplo, Faster <<http://www.faster-data.org>> (*Flexible Access to Statistics, Tables and Electronic Resources*) es un sistema de información estadística que pretende crear una herramienta flexible para acceder a las fuentes oficiales de información estadística, así como a otros datos estadísticos. Para ello plantea un entorno informativo que incorpora elementos de los sistemas de recuperación existentes, pero que basa la estructuración de la información en el desarrollo de metadatos. Este proyecto reúne grupos de investigación como el DSIG (*Statistics Domain Special Interest Group*) dentro del OMG (*Object Management Group*) y de la Iniciativa de Documentación de Datos (DDI), que se funda en el desarrollo de DTDs de XML para las ciencias sociales. Tratan de incorporar los estándares emergentes de metadatos como RDF y XMI, en aras a crear un modelo extensible y potente de metadatos que facilite la descripción de las particularidades de la información estadística como tablas, series temporales, etc. Para más información sobre metadatos estadísticos, *Vid.* <http://www.faster-data.org/Metadata/links/index.htm>

forma semejante, los Sistemas de Gestión de Bases de Datos (SGBD) contienen tradicionalmente, como apuntamos en el apartado anterior, un diccionario de datos que almacena, precisamente, metadatos para proteger la integridad de la información.

Finalmente, en una línea profética —que no refrendamos totalmente, en el sentido de que el límite entre el presente y el futuro, hablando de estos temas, es puramente anecdótico— Dempsey y Heery, opinan que *se espera que los futuros sistemas en red se basen en objetos distribuidos. Las aplicaciones únicas se disgregarán en componentes. Las nuevas aplicaciones se construirán a partir de múltiples componentes individuales especializados para tareas determinadas, que se reunirán cómo y cuándo sea necesario*²⁹, siendo los metadatos un aspecto relevante en esta nueva (actual y emergente) arquitectura de gestión e interacción entre objetos distribuidos. Con esta previsión para los sistemas de documentación del futuro, completan estos autores su aproximación a lo que pueden parecer metadatos en el contexto amplio y genérico de la gestión de información. Todos estos ejemplos tienen una relación directa con la visión bipartita que plantharemos en este acercamiento a la extensión y diferenciación del término metadatos.

Volviendo a la definición genérica, y ambigua, de *datos sobre los datos* y a los distintos dominios susceptibles de aplicarlos, nos encontramos con distintas interpretaciones del término metadatos que podríamos dividir en dos tendencias fundamentales. Por un lado la tendencia del "todo metadatos" —que coincidiría, en cierto sentido, con la caracterización a través de ejemplos de Dempsey y Heery, a la que nos hemos referido anteriormente—, y por otro, la interpretación exhaustiva que restringe el concepto de metadatos al contexto de la información electrónica distribuida a través de la Web que entiende los documentos como objetos, es decir,

²⁹ [...] *the future network system will be based on distributed objects. Monolithic applications will be disaggregated into components. New applications will be built from multiple autonomous components specialised for particular task, put together as and when required* [...] Lorcan Dempsey and Rachel Heery. Metadata: A Current View..., p. 147.

"metadatos en sentido estricto", aplicados a la WWW. No se trata de una clasificación categórica del concepto de metadatos, ya que en muchos casos, los mismos autores reconocen ambas acepciones. No obstante, esta división de posturas que exponemos a continuación, y su presentación en forma de dicotomía, simplifica nuestro objetivo de aportar una visión profunda del concepto en cuestión.

Dentro de la primera tendencia, que hemos denominado, de manera elemental pero significativa, "**todo metadatos**", encontramos afirmaciones como las siguientes:

Metadatos son datos asociados con objetos que liberan a sus usuarios potenciales de tener que poseer un conocimiento avanzado sobre su existencia o características (Dempsey y Heery³⁰).

Metadatos son datos sobre datos, y por lo tanto proporcionan una información básica como el autor de un trabajo, la fecha de creación, referencias a otros trabajos relacionados, etc. [...]. Los metadatos existen para casi todos los objetos o grupos de objetos [de información] imaginables, independientemente de que estén almacenados en formato electrónico o no (Miller³¹).

En realidad, los metadatos no son otra cosa que datos sobre los datos; un registro de un catálogo son metadatos; la cabecera TEI, también, o cualquier otra forma de descripción. Podríamos llamarlos catalogación, pero para algunas personas este término tiene demasiadas connotaciones, como las AACR y el USMARC. De esta forma, en cierto sentido, se trata de una

³⁰ Lorcan Dempsey and Rachel Heery. *A Review of Metadata: A Survey of Current Resource Description Formats* [documento RTF]. Bath: University, UKOLN Metadata Group, 21 de marzo de 1997 (versión 1.0), rev. 10 de junio de 1998. Disponible en: <http://www.ukoln.ac.uk/metadata/desire/overview/overview.rtf> , p. 5 (consultado el 30 de junio de 1998). *Cfr. infr.* Nota 37.

³¹ Paul Miller. *Metadata for the Masses* [documento HTML]. *Ariadne*, 11-September-1996, issue 5. Disponible en: <http://www.ariadne.ac.uk/issue5/metadata-mases> (consultado el 18 de junio de 1998).

situación del tipo 'you call it corn, we call it maize'³², pero metadatos es un término neutro bueno que cubre todas las bases (Caplan).

El conocido registro de catálogo de biblioteca se podría describir como metadatos, ya que dicho registro son datos sobre datos. De forma similar los registros de una base de datos realizada por un servicio de indización y resumen, son también metadatos (Rachel Heery³³).

Esta primera tendencia conceptual, como se deduce de las definiciones que hemos destacado, coincide, en esencia, con la finalidad de la práctica bibliotecaria y de los sistemas de organización y representación del conocimiento, que siempre han contado con estándares para describir la información y recuperarla posteriormente, de tal forma que los metadatos, llamados así o de cualquier otra manera, han existido

³² Caplan recurre a una metáfora en inglés: *you call it corn, we call it maize* (vosotros lo llamáis palomitas, nosotros, maíz) para referirse a que catalogación y metadatos son esencialmente lo mismo pero con distinto nombre. Priscilla Caplan. *You Call It Corn, We Call It Syntax-Independent Metadata for Document-Like Objects* [documento HTML]. *The Public-Access Computer Systems Review*, 1995, vol. 6, nº 4. Disponible en: <http://info.lib.uh.edu/pr/v6/n4/capl6n4.html> (consultado el 18 de junio de 1998). Es importante destacar que, a diferencia de lo que decían Dempsey o Gorman (*Cfr. sup. Notas 13 y 14*), para esta autora, "metadatos" es un término *bueno* y *neutro* que permite designar también la catalogación.

Otros autores, en esta misma línea, hablan de "una vieja respuesta a un viejo problema" (*metadata is really an age-old answer to an age-old problem*). Marty Lucas. *Demystifying Metadata* [documento HTML]. *Mappa Mundi*, April 2000. Disponible en: <http://mappa.mundi.net/trip-m/metadata> (consultado el 31 de octubre de 2000); o de "vino viejo en odres nuevos". Stefan Gradmann. *Cataloguing vs. Metadata: Old Wine in New Bottles?* [documento HTML]. En: *IFLA General Conference (64. 1998. Amsterdam)*. París: IFLA, INIST, rev. 23 de junio de 1998. Disponible en: <http://ifla.inist.fr/ifla/IV/ifla64/007-126e.htm> (consultado el 6 de julio de 1998).

³³ Rachel Heery. *Review of Metadata Formats* [documento HTML]. Bath: University, UKOLN, maintained by Michael Day, rev. 28 de enero de 1998. Disponible en: <http://www.ukoln.ac.uk/metadata/review.html> (consultado el 18 de junio de 1998). Pre-publicación de: Rachel Heery. *Review of Metadata Formats*. *Program*, October 1996, vol. 30, nº 4, [p. 345]. A pesar de esta definición tan genérica de Heery, que hemos insertado en este trabajo dentro del grupo conceptual "todo metadatos", la autora continúa diciendo que: *sin embargo el término metadatos está siendo cada vez más utilizado en el mundo de la información para determinar los registros que se refieren a los recursos digitales accesibles a través de una red [...]*. Según esto, debemos decir que la definición de Heery, en ésta y otras publicaciones —*Vid. Bibliografía*—, se inserta más en la postura de "metadatos en sentido estricto", entendidos como registros de datos vinculados a la información de la Red.

siempre. No obstante, algunos miembros de la comunidad bibliotecaria ponen obstáculos a este nuevo término, alegando que los metadatos son simplemente catalogación o datos bibliográficos con otro nombre, como hemos señalado. A pesar de que estas afirmaciones son, en cierto sentido, ciertas, existen muchas razones, desde nuestro punto de vista, para adoptar esta nueva terminología. Huc, Thierry y Nonon-Latapie³⁴ sugieren una acertada reflexión diacrónica, que puede servirnos para dirimir esta controversia conceptual. Estos autores justifican que el concepto de metadatos surgió en sistemas de gestión de grandes colecciones de objetos (museos y bibliotecas, por ejemplo), sin embargo, actualmente el principal problema surge en el ámbito de las colecciones de objetos digitales. Teniendo en cuenta esto, a la dicotomía conceptual que hemos planteado para definir los metadatos, podemos aducir además el punto de vista de continuidad en el tiempo, es decir, que la línea definitoria del "todo metadatos" recogería los inicios de las herramientas de descripción y recuperación de información tradicionales (catálogos, etc.), donde los metadatos son cualquier resultado de una actividad catalográfica y/o de descripción bibliográfica; mientras que la definición exhaustiva —que hemos denominado "metadatos en sentido estricto", sobre la que reflexionaremos a continuación— implicaría el actual problema de la descripción de la información en Internet, o lo que se ha denominado también la *catalogación de Internet*³⁵.

³⁴ Claude Huc, Levoir Thierry and Michel Nonon-Latapie. Metadata: Models and Conceptual Limits [documento HTML]. En: *IEEE Metadata Conference (2. 1997. Maryland)*. IEEE Computer Society, 1997. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/chuc/chuc.html> (consultado el 18 de junio de 1998).

³⁵ Denominaciones como ésta, o similares (catalogación de la Red, catalogación de recursos electrónicos, etc.) aparecen con frecuencia en la bibliografía sobre el tema, por ejemplo: Roman S. Panchyshyn y France Bouthellier. Cataloguer le cyberspace: le défi des ressources électroniques. *Documentation et Bibliothèques*, juillet-septembre 1997, vol. 43, n° 3, p. 137-147. Consideramos empero, que esta forma de nominar la descripción de recursos de la WWW debe abandonarse a favor de los metadatos. *Cfr.* 4.1.

Con esta apreciación diacrónica demostramos además la continuidad en la propuesta de soluciones bibliotecarias, que deben seguir siempre la correlación: a nuevos soportes/necesidades informativas nuevos planteamientos/soluciones bibliotecarias, basadas éstas últimas en la práctica acumulada en descripción y recuperación de información.

La segunda tendencia conceptual, "**metadatos en sentido estricto**", se refiere de forma exhaustiva o restrictiva al contexto de la información electrónica en la Red. En esta línea podemos subrayar definiciones como éstas:

Los metadatos son datos asociados con objetos [DLO³⁶] que liberan a sus usuarios potenciales [personas o programas] de la necesidad de tener un conocimiento avanzado completo sobre su existencia o características [...] Los metadatos son conocimiento que permite a los usuarios, humanos o automatizados, comportarse de manera inteligente (Dempsey y Heery³⁷).

Los metadatos son información inteligible por ordenador sobre recursos Web u otras cosas (Tim Berners-Lee³⁸). A esta definición concisa pero tajante, el

³⁶ Los DLOs (*Document Like Object*) son documentos entendidos como objetos de información. Si no entendiéramos en ese contexto el significado que tiene la palabra *objeto* —demasiado amplia y ambigua, por otra parte, en español— en la definición de Dempsey y Heery, no podríamos considerarla dentro de esta corriente definitoria de los metadatos "en sentido estricto" vinculados exclusivamente a la información de la Red. El concepto de DLO se explicará profusamente al hablar del fundamento de la metainformación. *Vid. infr.* 3.3.

³⁷ Lorcan Dempsey and Rachel Heery. Metadata: A Current View... *Op. cit.*, p. 149. Si bien esta definición es en esencia, idéntica a la que dábamos de estos mismos autores en la corriente definitoria anterior de "todo metadatos" (*Cfr. supr.* Nota 30), en esta ocasión Dempsey y Heery, aunque mantienen la facultad de los metadatos de sustituir o anticipar el contenido de un objeto, gradúan el concepto, aludiendo a la interpretación humana o automatizada de los metadatos. Esta inclusión en la definición es lo que nos ha permitido situarla en este bloque de "metadatos en sentido estricto", además de poder matizarla con la información incluida entre corchetes: datos asociados a [documentos entendidos como objetos de información] y datos que pueden ser interpretados tanto por [personas como por programas] que, como explicaremos, son dos de las constantes en esta nueva tendencia conceptual.

³⁸ *Metadata is machine understandable information about Web resources or other things.* Tim Berners-Lee. Metadata Architecture: Document, Metadata and Links [documento HTML]. *Design* [cont.]

teórico de la Web, añade además que la frase 'inteligible por ordenador' es clave, ya que reduce el concepto de metadatos a la información que usan los agentes de software.

[...] *los metadatos también difieren de los datos del catálogo tradicional en que la información de localización se encuentra en el registro de tal manera que permite obtener directamente el documento [...]; es decir, el registro puede contener información detallada de acceso y la dirección/es en la red (Heery³⁹).*

Los metadatos son 'quotinformation' sobre los datos. Describen un recurso de Internet: qué es, qué trata, dónde está, etc. Existen tres aspectos principales para el desarrollo de metadatos: descripción de recursos, producción de metadatos y uso de los metadatos (Iannella y Waugh⁴⁰).

El término usado normalmente para referirse a aquella información sobre objetos digitales (Goodchild⁴¹). O con palabras de Gill⁴², los metadatos son datos que sirven para describir grupos de datos a los que podríamos llamar objetos informáticos (Documentos como Objetos).

Todas estas definiciones que podemos considerar, en cierto sentido, tautológicas destacan tres aspectos fundamentales de los metadatos: la funcionalidad,

Issues. W3C, 6 de enero de 1997, rev. 6 de febrero de 1998. Disponible en: <http://www.w3.org/DesignIssues/Metadata.html> (consultado el 6 de julio de 1998).

³⁹ Rachel Heery. *Review of Metadata...* *Op. cit.*, <http://www.ukoln.ac.uk/metadata/review.html>

⁴⁰ Renato Iannella, and Andrew Waugh. Metadata: Enabling the Internet [documento HTML]. En: *Resource Discovery Unit*. Brisbane: University of Queensland, Distributed Systems Technologies, rev. 4 de febrero de 1997. Disponible: <http://www.dstc.edu.au/RDU/reports/CAUSE97> (consultado del 17 de julio de 1998). El término *quotinformation* puede referirse a información citada, en el sentido de que los metadatos son información ligada a la propia información. *Vid.* Glosario.

⁴¹ Michael F. Goodchild. *Alexandria Digital Library...* *Op. cit.*, http://www.alexandria.ucsb.edu/public-documents/metadata/metadata_ws.html

⁴² Tony Gill. Los metadatos y la World Wide Web. En: *Introducción a los metadatos: vías...* *Op. cit.*, p. 10.

el contexto y el interlocutor (hombre o máquina). A modo de resumen, la definición más exhaustiva que podemos subrayar en la línea de nuestra tesis, deriva de las ideas de Hudgins, Agnew y Brown, de Vellucci y del propio Consorcio Web (W3C), donde se ponen de manifiesto estos tres aspectos vinculados al término metadatos: la función identificadora y descriptiva de los metadatos, el contexto de la Red y la posibilidad de interpretación por la máquina: Los registros de metadatos se refieren, y generalmente sólo existen o tienen significado en relación con los documentos u objetos referenciados. El término metadatos alude a cualquier dato que ayuda en la identificación, descripción y localización de recursos electrónicos en red⁴³. Son pues, datos que describen los atributos de un recurso, caracterizan sus relaciones, soportan su recuperación, gestión y uso efectivo, y existen en un contexto estrictamente electrónico⁴⁴. A lo que se puede añadir el pensamiento corporativo del Consorcio Web, que asume la definición de Berners-Lee como información entendible por máquina para la Web, y precisa más concretamente el concepto de metadatos como información sobre la información, información etiquetada, catalogada y descriptiva, estructurada de tal forma que permita que las páginas Web se procesen y se busquen correctamente, en particular, por el ordenador⁴⁵. Estas ideas reflejan que el problema actual concomitante a los metadatos reside en contextos de información

⁴³ Jean Hudgins, Grace Agnew, Elizabeth Brown. *Getting Mileage out of Metadata: Applications for the Library*. Chicago: ALA, 1999, p. 1.

⁴⁴ Esta definición la encontramos en las siguientes contribuciones: Sherry L. Vellucci. Metadata and Authority... *Op. cit.*, p. 34; y Sherry L. Vellucci. Metadata. *Annual Review of Information Science and Technology (ARIST)*, 1998, vol. 33, p. 192.

⁴⁵ World Wide Web Consortium. *Metadata: Activity Statement* [documento HTML]. W3C, rev. 10 de octubre de 2000. Disponible en: <http://www.w3.org/Metadata/Activity.html> (consultado el 3 de enero de 2001). Esta definición, o pequeñas variantes de ella, es la más aceptada en general en la bibliografía profesional. Citando o no al consorcio Web, las empresas, instituciones o personas vinculadas al mundo de la metainformación asumen esta concepción. V. gr. la empresa Watchfire que comercializa el software para la gestión de metadatos Metabot reproduce esta definición en: *Metatags: They're not Just for Search Engines: White Paper* [documento PDF]. Kanata: Watchfire, March 2000. Disponible en: <http://www.watchfire.com/resources/metatagswhite.pdf>, p. 1 (consultado el 4 de abril de 2001).

protagonizados por elementos u objetos electrónicos, donde tanto datos como metadatos están incluidos en cadenas de bits y la representación de la información es doble, ya que puede estar determinada por datos o por metadatos, pudiendo acceder, tanto los usuarios como la máquina, a ambos.

Al igual que nosotros hemos hecho una división del concepto en dos tendencias o líneas definitorias, Burnett, Bor Ng y Park⁴⁶ distinguen también una doble aproximación a los metadatos: por un lado *el enfoque del control bibliográfico* cuyo origen está en la Biblioteconomía, y por otro *el enfoque de la gestión de datos*, cuyos orígenes están en la Informática. Esta doble visión corrobora también otra idea que sostenemos en este trabajo, que tanto el concepto de los metadatos como su aplicación tiene que entenderse desde esta doble perspectiva: la de la práctica acumulada en las bibliotecas, y la de la informatización de la gestión de la información.

La *aproximación del control bibliográfico* hace referencia a la larga tradición de las bibliotecas en el desarrollo de sistemas de información y en el establecimiento de reglas y estándares para organizar y acceder a los recursos de información almacenados en ellas; de igual forma y con el mismo fin (controlar y compartir la información), se están desarrollando estándares para afrontar la información electrónica⁴⁷. En este sentido podemos evocar, por ejemplo, la definición que da la

⁴⁶ Kathleen Burnett, Kwong Bor Ng, Soyeon Park. A Comparison of the Two Traditions of Metadata Development. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, n° 13, p. 1209-1217.

⁴⁷ Una muestra de posturas en torno a los metadatos desde esta perspectiva del control bibliotecario podemos apreciarla en las ponencias de la *Conferencia del Bicentenario del Control Bibliográfico*, celebrada en el 2000 en la Universidad de Campinas (Brasil). En tal evento se puede apreciar la aproximación a los metadatos de, por ejemplo, Michael Gorman, Priscila Caplan o Martin Dillon, partiendo de la tradición bibliotecaria y alimentando las nuevas necesidades de acceso a la información en la Web. Para acceder a todas las comunicaciones pronunciadas dicha conferencia, consultar: <http://www.unicamp.br/bc/congresso.html> (Vid. Bibliografía, para una descripción detallada de las contribuciones por autores).

IFLA⁴⁸ de metadatos como *cualquier dato utilizado para ayudar a la identificación, descripción y localización de recursos electrónicos en red*. Asimismo, dentro de este enfoque, el director de la *Alexandria Digital Library* (ADL)⁴⁹ enumera las características de los metadatos en el contexto bibliotecario tradicional adaptándolas al contexto digital:

- La organización de la metainformación en forma física se cambia por una organización de forma electrónica más flexible, pero más estructurada.
- La organización física de la información en un catálogo se sustituye por múltiples formas de organización lógica de los objetos de información.
- El hecho de tener los objetos de información en formato digital facilita también el uso de la tecnología digital para su extracción e identificación.

Esta aproximación desde el punto de vista del control bibliográfico corrobora nuestra visión de que el concepto de metadatos es un *continuum* que implica, como decíamos antes, nuevas soluciones bibliotecarias ante las nuevas formas de producción del conocimiento.

Por otro lado, el *enfoque de la gestión de datos* parte de la tradición de la comunidad informática en la gestión y organización de información a través de la creación de sistemas de almacenamiento y recuperación para datos textuales y

⁴⁸ [...] *any data used to aid the identification, description and location of networked electronic resources*. IFLA. *Digital libraries: Metadata Resources* [documento HTML]. The Hague: IFLA, 1995-2000, rev 3 de octubre de 2000. Disponible en: <http://www.ifla.org/ll/metadata.htm> (consultado el 10 de febrero de 2001).

⁴⁹ Terence R. Smith. *Meta-Information Environment of Digital Libraries* [documento HTML]. *D-Lib Magazine*, July-August 1996, vol. 2, n° 7/8. Disponible en: <http://www.dlib.org/dlib/july96/new/07smith.html> (consultado el 18 de junio de 1998). También disponible en: <http://www.alexandria.ucsb.edu/public-documents/dlib> (consultado el 6 de febrero de 2001). Retomaremos este planteamiento de Smith para caracterizar el concepto de biblioteca digital. *Cfr.* Capítulo 8, nota 14.

relacionales. Estos sistemas informáticos proporcionan funciones de seguridad e integridad que permiten compartir datos que, añadidas a las funciones de localización, identificación y recuperación de los sistemas de control bibliográfico, han tenido también un papel importante en el desarrollo de esquemas y modelos de metadatos. En esta visión de cariz más informático, podríamos encuadrar definiciones como esta: *los metadatos son información documentada a través de herramientas de tecnologías de la información que mejoran la comprensión, tanto técnica como comercial, de los datos y de los procesos relacionados con ellos*⁵⁰. Los metadatos se entienden pues, como *información adicional necesaria para que los datos sean útiles*⁵¹, es decir, que facilitan la gestión, el acceso o el análisis de datos

⁵⁰ Robert S. Seiner. *Questions Metadata can Answer* [documento PDF]. Computer Associates Products, CAI, 14 de julio de 2000. Disponible en: http://www.cai.com/products/decisionbase/questions_metadata_can_answer.pdf (consultado el 24 de julio de 2000). Esta misma definición la citamos en un trabajo anterior, en el que tratamos de poner de relieve la utilidad y necesidad de los metadatos orientados al contenido, en entornos finitos de información, concretamente en ese caso, en las Intranets corporativas. Eva M^a Méndez Rodríguez. Metadatos y tesauros: aplicación XML/RDF a los sistemas de organización del conocimiento en Intranets. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao). La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p. 213.

⁵¹ Esta es una afirmación recurrente en foros sobre metadatos en el ámbito informático, p. ej., el segundo congreso del Institute of Electrical and Electronics Engineers. *IEEE Metadata Conference (2. 1997. Maryland)*. IEEE Computer Society, 1997. Disponible en: http://www.llnl.gov/liv_comp/metadata/md97.html (consultado el 5 de febrero de 2001). Definiciones semejantes se advierten en autores como Clive Finkelstein, gurú de la ingeniería de la información, que plantean su visión de los metadatos desde la perspectiva de la arquitectura de datos a nivel organizacional, como queda patente, en los *simposia* que celebra anualmente DAMA (*Data Management Association* <<http://www.dama.org>>), o en el ámbito de la informática decisional y de gestión.

No obstante, esta concepción de los metadatos —a la que nos adscribimos (*Cfr. infr.*)— está suficientemente aceptada como vemos también en: Martin Grötschel, Joachim Lügger. *Scientific Information Systems and Metadata* [documento HTML]. *ZIB Report SC-98-26*. Berlin: Konrad-Zuse-Zentrum für Informationstechnik, 26 de octubre 1998. Disponible en: <http://elib.zib.de/ftp/pub/UserHome/Luegger/Dresden/Metadata.htm> (consultado el 22 de agosto de 2000); o en definiciones como [...] *meta data as any information required to make other data useful*. Bill Prothman. *Meta Data: Managing Needles in the Proverbial Haystacks* [documento PDF]. *IEEE Potentials*, february-march 2000. Disponible en: <http://www.ieee.org/membership/students/potentials/febmar2000/pdf/pages20.pdf>, p. 20 (consultado el 13 de julio de 2000).

que, como veremos, serán la base de las aplicaciones que se le da, en el contexto de Internet, a la metainformación.

Del doble enfoque del concepto en Burnett, Bor Ng y Park, también trasciende una definición que unifica las dos tendencias conceptuales señaladas por ellos:

[...] *los metadatos son datos que caracterizan datos fuente, describen sus relaciones, y soportan su recuperación y uso efectivo*⁵²,

y que pone de manifiesto otro aspecto que también apoyamos en esta tesis: que el concepto de metadatos está teñido por un componente de gestión y procesamiento bibliotecario a la par que por un componente de gestión y procesamiento informático y que su función principal es la recuperación de información.

Por su parte, Dovey⁵³ también hace una interesante segmentación del concepto de metadatos identificando tres corrientes principales a las que denomina "escuelas": *la escuela de la catalogación, la escuela estructuralista y la escuela de la estructura de datos.*

La primera de estas escuelas, la de los catalogadores, es la más antigua y quizás la más amplia; coincide básicamente con la tendencia conceptual que nosotros hemos denominado "todo metadatos", que incluye tanto a los catalogadores, como a los agentes que asignan metadatos, junto con un pequeño grupo de profesionales que entienden que ambas actividades son, en esencia, lo mismo. Dentro de esta escuela, los catalogadores, obviamente proceden del mundo bibliotecario y han estado generando metadatos prácticamente desde que existen las bibliotecas. Pero recientemente han descubierto que lo que han estado haciendo durante tanto tiempo

⁵²... *metadata is data that characterizes source data, describes their relationships, and supports the discovery and effective use of source data.* Kathleen Burnett, Kwong Bor Ng, Soyeon Park. *Op. cit.*, p. 1212.

⁵³ Matthew J. Dovey. "Stuff" about "Stuff"... *Op. cit.*, p. 6-13.

era efectivamente crear metadatos⁵⁴, primeramente en forma de fichas de catálogo y, a partir de finales de los 60, en formato MARC, que comenzó siendo un formato de transferencia e intercambio de registros y que se ha usado para la entrada, presentación, e incluso almacenamiento de información bibliográfica.

El otro grupo que corresponde a esta escuela es, paradójicamente, el de los informáticos que, a tenor del crecimiento exponencial de la World Wide Web, están preocupados por una estructura que haga a la información más manejable, es decir, están preocupados por construir metadatos para catalogar esa información. Los informáticos, a diferencia de los catalogadores, tienen experiencia en el desarrollo de protocolos y de software, pero esto sólo no es suficiente para diseñar estándares para la gestión del conocimiento; por ello, a la hora de desarrollar metadatos se concentran en el formato y la sintaxis, ignorando el contenido o la semántica, y piensan que cuando crean formatos de metainformación y reglas están catalogando, de la misma manera que los bibliotecarios cuando catalogan creen que están creando metadatos. A esta escuela o forma de entender los metadatos corresponde, en cierto sentido, la concepción emanada de formatos como el Dublin Core y el RDF (*Resource Description Framework*), de los cuales hablaremos más detenidamente en el capítulo 5 dedicado a los modelos o formatos de metainformación.

La segunda escuela que identifica Dovey, es la que denomina "estructuralista" y está constituida básicamente por los creadores de los lenguajes de marcado. Al igual que ocurría con la de los catalogadores, se trata de una vieja escuela, que tiene sus orígenes en una disciplina todavía más antigua, anterior incluso a los ordenadores y al despliegue informático, como es el conocimiento humanístico y más concretamente, el análisis textual.

⁵⁴ Con la misma ironía, aunque con un planteamiento inverso, Brisson comienza su artículo —al que nos hemos referido en este trabajo, *Cfr.* Capítulo 2, nota 1— aludiendo a que, a raíz de la proliferación y de, por qué no, la moda de los metadatos, el mundo descubre la catalogación. Roger Brisson. *The World Discovers Cataloging a Conceptual Introduction to Digital Libraries, Metadata... Op. cit.*, p. 3.

El estructuralismo es la teoría y método científico, utilizado en diferentes ciencias⁵⁵, que considera un conjunto de datos como una estructura o sistema de interrelaciones. Dovey, sin embargo, habla de una escuela estructuralista refiriéndose a la *invención* del marcado textual como *la habilidad de embellecer un fichero de texto ASCII con comentarios adicionales y señales en lo que se refiere a sus estructuras subyacentes e interpretativas*⁵⁶. A esta escuela pertenecen la forma de entender los metadatos en estándares o modelos como SGML (*Standard Generalized Mark-up Language - ISO 8879*), o TEI (*Text Encoding Initiative*) que constituye, como veremos, una aplicación de SGML (XML) para el marcado de textos y sobre todo, cada vez más, el marcado flexible XML hacia el que evolucionan todas las perspectivas de metadatos basadas inicialmente en SGML. Además una de las mayores aportaciones de esta escuela es, la definición, lacónica pero explícita, de metadatos como: *datos estructurados sobre los datos*⁵⁷.

Desde nuestro punto de vista, Matthew Dovey denomina a esta tendencia en la aprehensión de la organización de la información, "estructuralista", por un parte evidente, porque los estándares y modelos que identifica dentro de ella, están basados en la estructura del texto, y por otra, porque se puede hacer un trasvase del

⁵⁵ En general, el estructuralismo es la tendencia filosófica que surgió en la década de los 60, y que pretendía reflejar un nuevo "estilo de pensar". Esta corriente reúne autores muy diferentes, que se expresan en los más diversos campos de las ciencias humanas, como p. ej. Lévi-Strauss en antropología, Barthes en crítica literaria, Foucault en la investigación historiográfica, o en corrientes filosóficas específicas como el marxismo de Althusser. Sin embargo, el estructuralismo obtiene sus instrumentos de análisis de la lingüística. De hecho, un punto de referencia común a los distintos desarrollos del estructuralismo ha sido siempre la obra de Ferdinand de Saussure. *Curso de lingüística general* (1915). Buenos Aires: Alianza, 1989.

⁵⁶ Matthew J. Dovey. "Stuff" about "Stuff"... *Op. cit.*, p. 10.

⁵⁷ *Structured data about data*, esta definición de los metadatos, sencilla y palmaria a la vez, la encontramos en muchos autores, v. gr. en Katherine Burnett, Kwong Bor Ng, Soyeon Park. *Op. cit.*, p. 1209 y 1212; o en Renato Iannella. *Mostly Metadata: A Bit Smarter Technology* [documento HTML]. En: *Resource Discovery Unit*. Brisbane: University of Queensland, Distributed Systems Technologies, 1997. Disponible: <http://www.dstc.edu.au/RDU/reports/VALA1998> (consultado el 21 de julio de 1998).

estructuralismo lingüístico a los lenguajes informáticos. El estructuralismo saussuriano supuso una revolución epistemológica que rompía con la lingüística tradicional, la cual, consideraba al lenguaje como una mera nomenclatura, donde cada signo tendría una identidad propia, y la tarea de la lingüística era puramente descriptiva. De la misma forma, el estructuralismo de la información, es decir, el marcado de documentos, supone también una revolución en la definición de la sintaxis y de las reglas para definir las etiquetas que indican las estructuras y cómo dichas etiquetas deben usarse en un documento.

Desde hace poco tiempo, esta escuela estructuralista se centra en, el relativamente recién llegado a la escena de los lenguajes de marcado, XML (*eXtensible Mark-up Language*) que pone de relieve muchos de los temas recurrentes en el ámbito de los metadatos, sobre todo, el problema de la separación del contenido, de la presentación en los documentos electrónicos (*Cfr.* capítulo 5).

XML nos conduce a la última de las escuelas que distingue Dovey, una nueva escuela que ve al XML como un lenguaje universal para definir estructuras de datos. El origen de esta escuela, que este autor denomina *de estructura de datos*, está basado esencialmente en el crecimiento de Internet. La escuela de la estructura de datos, se apoya, como anticipábamos antes, en la separación de los datos, de la presentación utilizando XML y XSL (*eXtensible Style Sheet Language*). XML se muestra además, como un formato adecuado para el intercambio entre bases de datos, es decir, se destaca la posibilidad de utilizar XML para definir e intercambiar tanto la estructura semántica de las bases de datos, como el contenido. Pero además, de manera progresiva, se ve como un estándar para editar, manipular y almacenar datos, de la misma forma que el formato MARC se convirtió en un formato de intercambio en el mundo de la información bibliográfica.

Esta última escuela, reciente e innovadora, plantea algo que no es estrictamente nuevo en el mundo de la normalización y el intercambio de información: un lenguaje universal para anotar estructuras de datos. Esta misma idea se proponía ya en un estándar ISO de principios de los 90's, denominado ASN.1⁵⁸. Sin embargo la nueva concepción, también estructuralista de los metadatos, fundada en XML, abre una nueva época en la descripción de información a la que nos adscribimos, donde destacan posturas como la de Berners-Lee, Ianella, etc. y sobre la que tendremos oportunidad de hablar profusamente a lo largo de esta tesis doctoral.

3.3. Fundamento de la metainformación: el Documento como Objeto

De todo el conjunto de definiciones, tendencias conceptuales, escuelas, etc. discutidas en el apartado anterior, debemos destacar la conclusión de Gill⁵⁹ que opina que los metadatos son simplemente *datos significativos que representan otros objetos de datos discretos [...] descripciones estructuradas de un objeto informático* (de un Documento como Objeto). Sin embargo, el problema filosófico y el fundamento del concepto de metadatos, desde esta posición, residen en definir qué es

⁵⁸ ISO 8824 — Information Processing Systems - Open Systems Interconnection - Specification of Abstract Syntax Notation One (ASN.1), 1990. La importancia o influjo de la norma ASN.1 se puede ver en la labor de la oficina australiana del W3C, el DSTC (Distributed Systems Technology Centre), que está desarrollando las reglas que describen cómo codificar datos especificados según la norma ASN.1 como documentos XML, denominadas *XML Encoding Rules (XER)*. Hoylen Sue. XER - A Bridge between ASN.1 and XML [documento HTML]. Brisbane: University of Queensland, Distributed Systems Technologies Centre, 12 de enero de 1998, rev. 23 de junio de 1999. Disponible en: <http://www.dstc.edu.au/Research/Projects/xer/xer-2.html> (consultado el 12 de abril de 2001).

⁵⁹ Tony Gill. Los metadatos... *Op. cit.*, p. 10. La reafirmación de la concepción del documento digital como objeto de información así como una actualización de todos los planteamientos publicados originalmente en 1998, los encontramos en la revisión electrónica esta monografía, una de las primeras impresas sobre metadatos, coordinada por Murtha Baca. *Vid.* Tony Gill. Metadata and the World Wide Web. En: *Introduction to Metadata: pathways to Digital Information*. [version 2.0] [documento HTML]. Murtha Baca, ed. Los Angeles: Getty Research Institute, 1 de agosto 2000. Disponible en: http://www.getty.edu/research/institute/standards/intrometadata/2_articles/gill/content.html#introduction (consultado el 8 de agosto de 2000).

exactamente un objeto de datos o, más precisamente, qué es un objeto informático o qué es un Documento entendido como Objeto en el contexto de la información distribuida.

Aunque el planeamiento de la información como objeto no es nuevo, ya que algunos autores como Buckland la han tratado, con mucha anterioridad, de forma semejante (*information-as-thing*), lo realmente provocador en el contexto de los metadatos es aislar o definir ese objeto informático al que se refiere Gill, en un entorno WWW donde la información parece más un fenómeno, que un conjunto de documentos aislables. Esto es, el concepto de objeto de información pertenece al dominio informativo de la Web y es mucho más abstracto que cualquier identificación de la información como "cosa"; además a esto hay que añadir la vaguedad y amplitud del término "objeto" en nuestro idioma.

Esencialmente, cualquier producto intelectual en forma de Objeto Digital consiste definición estandarizada de identificador único (URI) y en un cuerpo monográfico de contenido que tiene la intención de publicarse y accederse electrónicamente⁶⁰. Según esto y partiendo del modelo de referencia del libro impreso, la relación entre un objeto digital de datos y los metadatos que lo describen es, desde un punto de vista puramente funcional, la misma que existe entre un libro y su registro catalográfico en una biblioteca.

Las bibliotecas tradicionales se caracterizan por tres aspectos organizacionales que son los que aportan la base para el acceso a la información: 1) la organización en objetos de información físicos (libros, revistas, etc.); 2) la organización de la

⁶⁰ Esta interpretación de Objeto Digital está basada en una definición que la AAP (*Association of American Publishers*) da acerca de libro electrónico (*ebook*). Citada en: Mark Bide & Associates. *Standards for Electronic Publishing: An Overview* [documento PDF]. NEDLIB Project. La Hogue: National Library of Netherlands, Koninklijke Bibliotheek, agosto de 2000. Disponible en: <http://www.kb.nl/coop/nedlib/results/e-publishingstandards.pdf>, p. 8 (consultado el 24 de febrero de 2001).

colección en objetos de información según su materia, autor, etc., y 3) un entorno informativo organizado que facilita el acceso directo a esos objetos de información, y un acceso indirecto al contenido de los mismos, gracias al bibliotecario y a su trabajo técnico que se materializa en catálogos y en la propia organización física, por materias, de los libros en las estanterías. Este entorno organizado es lo que Smith denomina: *el contexto de meta-información de una biblioteca*⁶¹. De tal forma que — opina Oddy⁶²— el aspecto más importante de la estructura organizacional de una biblioteca es que es visible, además de lógica y rápidamente comprensible.

Sin embargo, en las bibliotecas digitales y en general, en la información distribuida en Internet, las colecciones están formadas por texto, imagen, audio y vídeo en formatos de codificación numérica, esto es, por documentos entendidos como objetos —*Documents Like Objects* (DLO⁶³) o *Document-like Information Objects* (DIO⁶⁴)—, y el acceso a la información se basa en la tecnología informática. Así, la organización en objetos de información se convierte en una organización de objetos de información lógicos y el contexto de la metainformación cambia, en el sentido de que la organización de los objetos digitales no necesitan una interacción constante entre el usuario y el profesional de la información, de tal forma que desaparecen los intermediarios en el ciclo de vida del documento.

⁶¹ Terence R. Smith. *Op. cit.*, <http://www.dlib.org/dlib/july96/new/07smith.html>

⁶² Pat Oddy. *Op. cit.*, p. 25.

⁶³ El acrónimo DLO, y el concepto aparejado a él (*Document Like Object*), surge, como veremos, en el seno del desarrollo del Dublin Core. *Vid.* Stuart Weibel and Eric Miller. Image Description on the Internet: A Summary of the CNI/OCLC Image Metadata Workshop, September 24-25, 1996, Dublin, Ohio [documento HTML]. *D-Lib Magazine*, January 1997, vol. 3, n° 1. Disponible en: <http://www.dlib.org/dlib/january97/oclc/01weibel.html> (consultado el 18 de junio de 1998).

⁶⁴ McKemmish, Cunningham y Parer, matizan incluyendo la palabra "información" en la definición. Sue McKemmish, Adrian Cunningham and Dagmar Parer. *Metadata Mania* [documento HTML]. Melbourne: Monash University, School of Information Management & Systems, Records Continuum Research Group, rev. 5 de enero de 2000. Disponible en: <http://www.sims.monash.edu.au/rcrg/publications/recordkeepingmetadata/sm01.html> (consultado el 31 de agosto de 2000).

Una vez más no existe un consenso ni una definición clara para los documentos electrónicos de Internet entendidos como objetos (DLOs o DIOs)⁶⁵. Podríamos decir que un objeto de información es, en términos generales, un grupo de elementos digitales que un ordenador puede manipular como si fuese un objeto único independientemente del formato que tengan. La abstracción de DLO (Documento como Objeto) es una cuestión clave en el ámbito de la información distribuida en la Red y el fundamento de los metadatos necesarios para caracterizarla y recuperarla. Es un concepto emanado del primer *workshop* del Dublin Core para diferenciar nociones individuales que constituyen un objeto discreto digno de una descripción aparte⁶⁶. Desde entonces, la expresión *Document Like Object/s* está ampliamente reconocida en la literatura sobre metadatos para referirse a los documentos de Internet, que pueden estar formados por texto, imagen, audio o vídeo e hipertexto que invoque otros DLOs. En otros contextos también se utiliza el acrónimo IBOs (*Information Bearing Objects*) para referirse a los objetos que conllevan información independientemente del formato de ésta. Existe incluso una categorización de estos elementos, en forma de tesoro, para la tipificación de los objetos digitales y su ulterior recuperación o exclusión en las búsquedas⁶⁷; e incluso en algunos de los

⁶⁵ Algunos autores como Gill, destacan esta falta de concreción sobre el concepto denunciando lo vago y elusivo del término DLO que sólo se había definido, hasta entonces (1997), a través de ejemplos. Por ello, se refiere a la noción de *Document Like Object* en un tono irónico, recogiendo las palabras de un comentarista que señaló que *los metadatos para la recuperación de información en Internet eran un GLO, o Grial como Objeto (Grail-Like Object)*. Tony Gill. The Dublin Core Metadata Element Set: Useful Tool or the Emperor's Newest Clothes? *Spectra*, Winter 1997, p. 41.

⁶⁶ Stuart Weibel and Eric Miller. *Op. cit.*, <http://www.dlib.org/dlib/january97/oclc/01weibel.html>

⁶⁷ Para categorizar los tipos de DLOs (o *Information Bearing Objects*) la *Alexandria Digital Library* ha desarrollado un tesoro que asista para completar el contenido de las etiquetas de metadatos relativas al *tipo de documento* en los distintos formatos. *Draft Thesaurus of Information Object Type Terminology* [documento HTML]. Alexandria Digital Library. Santa Bárbara: University of California, 6 de agosto de 1998. Disponible en: http://www.alexandria.ucsb.edu/~lhill/MultiTes_files/objtype.tes/index.htm (consultado el 28 de agosto de 1998); 12 de mayo de 2000, http://alexandria.sdc.ucsb.edu/~lhill/MultiTes_files/objtype.tes/tagged.txt (consultado el 10 de febrero de 2001).

primeros proyectos de organización de recursos electrónicos encontramos la expresión *Electronic Information Objects* (EOI⁶⁸), con un significado parecido: objetos de información electrónica.

Frente a los DLOs están los no-DLOs, entendiendo como tales, por ejemplo, experiencias virtuales, bases de datos que generan "documentos como resultados" (*Document-Like Outputs*), o aplicaciones interactivas que pueden tener un contenido diferente según qué usuarios las utilicen. El concepto de DLO, es fundamental para el establecimiento de sistemas de recuperación de información en Internet basados en metadatos, entendidos como verdaderos sistemas y no sólo como meros conjuntos de textos, imágenes, etc.

Los catálogos, en sentido amplio, constituidos por descripciones perfectamente estructuradas de objetos de información, pueden asociarse a cualquier colección de objetos que necesite organizarse (un acervo de libros, una colección de CDs particular, o un conjunto de piezas de un museo). De igual forma, en el contexto automatizado o electrónico, cada registro descriptivo es un intento de capturar de forma concisa y útil las características clave, la esencia —formal y/o de contenido— del objeto de información.

En la escena digital, el ciclo de vida de un documento entendido como objeto (DLO), pasa por cinco etapas o fases señaladas, entre otros, por Gilliland-Swetland⁶⁹:

⁶⁸ Dennis Nicholson, et al. *Cataloguing the Internet: CATRIONA Feasibility Study. Report To The British Library Research & Development Department* [documento HTML]. The British Library Board, BUBL, 1995. Disponible en: <http://bubl.ac.uk/org/catriona/cat1rep.htm> (consultado el 22 de abril de 2001). Cfr. 4.1. donde hablaremos expresamente del proyecto CATRIONA.

⁶⁹ Anne J. Gilliland-Swetland. *Op. cit.*, p. 5. En este aspecto hemos utilizado la versión original en inglés actualizada en el 2000 de dicha obra editada por M. Baca. *Introduction to Metadata: pathways...* *Op. cit.*, http://www.getty.edu/gri/standard/intrometadata/2_articles/swetland/content.htm#categorizing

1. Etapa de creación: Los documentos entendidos como objetos, entran en un sistema de información digital, bien creándose originalmente de forma digital (v. gr., escribir un texto en formato HTML, o realizar una fotografía con una cámara digital), bien convirtiendo a formatos digitales objetos preexistentes en forma de átomos (por ejemplo, escanear un texto y almacenarlo en formato electrónico con la ayuda de un programa de reconocimiento óptico de caracteres —OCR— o digitalizar una imagen con ayuda de un escáner). Incluso, en muchas ocasiones, se crean múltiples versiones de un mismo objeto con finalidades de conservación, investigación o de difusión. En cualquier caso, el creador del objeto digital debe incluir metadatos de tipo descriptivo para identificar los recursos de información, así como metadatos para su gestión y administración.
2. Fase de organización: los objetos se organizan manual o automáticamente en la estructura del sistema de información digital y pueden introducirse metadatos adicionales a través de un proceso de indización, catalogación o registro.
3. Etapa de búsqueda: en esta fase los documentos como objetos de información, están sujetos a un proceso de búsqueda y recuperación por parte de los usuarios potenciales. Por ello, el sistema informático crea metadatos para producir algoritmos de búsqueda, y proporcionar la eficacia del sistema en términos de almacenamiento y recuperación.
4. Fase de uso: los recursos de información (DLOs), una vez recuperados, se utilizarán, reproducirán y modificarán. Así, se pueden crear metadatos como notas de uso, derechos de autor o control de versiones de los datos.
5. En la última fase, de conservación y disposición, los objetos de información se someten a procesos como actualización, migración, comprobación de la integridad para asegurar su disponibilidad en el futuro, de igual forma que se pueden eliminar los DLOs que están inactivos o que ya no son necesarios. En este sentido, los metadatos pueden documentar ambas actividades de conservación y de disposición o disponibilidad.

Según esta interpretación de las fases por las que pasan los DLOs, entendiendo esta unidad informativa digital como fundamento de los metadatos, podemos completar el concepto de estos últimos, añadiendo, a todas las definiciones consignadas en el apartado anterior, que: los metadatos pueden incluir datos asociados con cualquier sistema de información u objeto de información con la finalidad de realizar una descripción, facilitar la administración de la información, reflejar las condiciones legales de uso, su funcionalidad técnica, y sus particularidades de conservación o continuidad en la Red.

La Biblioteca del Congreso de los Estados Unidos⁷⁰ propone además una jerarquía para poner en orden la diversidad de objetos digitales, DLOs o simplemente, objetos informáticos de la Red, de tal forma que los elementos de metadatos asociados al concepto de DLO, puedan aplicarse a múltiples y variados niveles:

- Nivel de colección (*set level*): este nivel de metadatos se aplica a lo que normalmente se conoce como colección digital, formada por conjuntos que agrupan ítems digitales por el tipo de contenido original (texto, documento visual o audio, etc.), y por la responsabilidad de la custodia del DLO, así como por la colección. Una colección puede determinarse por series de archivo o por materia bibliográfica. Algunas están contenidas en un solo conjunto, y otras se forman a partir de varios. Este nivel de metadatos se aplica a todos los conjuntos dentro de una colección digital independientemente de su contenido y/o responsabilidad.
- Nivel de conjunto (*aggregate level*): un conjunto organiza DLOs por tipo de documento digital y por la responsabilidad de la custodia del mismo; a este nivel

⁷⁰ *Library of Congress Digital Repository Development: Core Metadata Elements* [documento HTML]. Washington, DC.: Library of Congress, 18 de mayo de 2000. Disponible en: <http://lcweb.loc.gov/standards/metadata.html> (consultado el 28 de agosto de 2000).

se puede considerar también una colección digital. Los metadatos a nivel conjunto se aplican a todos los objetos primarios que forman parte del mismo.

- Nivel de objeto primario. Un objeto primario —según la *Library of Congress*— es el ítem identificado como un todo coherente, y suelen ser el equivalente digital a un ítem físico, a una unidad bibliográfica, en una biblioteca tradicional, es decir, a un libro, una grabación sonora, una videograbación, una partitura musical, un mapa, etc. Los metadatos se aplican a todos los objetos intermediarios y terminales de un objeto primario concreto.
- Nivel de objeto intermediario. El objeto intermediario es una visión o un componente de un objeto primario. Por ejemplo, si tomamos como objeto primario, un libro electrónico en la Web, que contenga imágenes y texto, se puede decir que tiene dos objetos intermedios. Los metadatos asociados a este nivel permitirán la recogida de archivos digitales y metadatos para la creación de presentaciones.
- Nivel de objeto terminal. El objeto terminal es el fichero o ficheros de contenido digital que forman el objeto. Hay al menos un objeto terminal para cada objeto. Los metadatos a este nivel son sobre todo estructurales y proporcionan los atributos digitales de cada fichero, tales como tamaño, extensión, etc.

En cualquier caso, creemos que la estructura de un registro descriptivo debe tener como objetivo fundamental la búsqueda, ordenación y recuperación de todos estos niveles de información digital, con el fin de facilitar la gestión y el uso de la colección; del mismo modo que deben llevarse a cabo las tareas de organización para alcanzar un grado de homogeneidad y estructuración que hagan recuperables y útiles los DLOs.

De toda la discusión conceptual que hemos expuesto hasta aquí, tanto en lo relativo al término metadatos, como a su fundamento (el Documento como Objeto o DLO), podemos urdir una definición que se ajuste al alcance de este trabajo de investigación y que resuma todas las ideas reveladas hasta ahora: Los metadatos son

pues, informaciones que hacen útiles los datos. Están destinados a ordenar y describir la información contenida en un documento entendido como objeto (DLO), de tal forma que se erigen como reveladores tanto de la descripción formal, como del análisis de contenido, en aras a mejorar el acceso a los objetos de información de la Red. Desde nuestro punto de vista, no son más que estructuras de organización de la información, legibles por máquina como señala Berners-Lee, cuya finalidad es incrementar la relevancia en los resultados de búsqueda en Internet, o sea, hacer útiles los datos. Y esa forma de hacer útiles los datos será diferente según las necesidades concretas y por ende, según la aplicación que se les otorgue.

3.4. Aplicación de los metadatos

Para completar la definición, fundamento y caracterización de la metainformación, nos resta abordar el tema de para qué sirve, es decir, reflexionar profusamente en torno a la función o aplicación de los metadatos. En primer lugar, podemos hacer una diferenciación con las aplicaciones de los metadatos en la gestión de la información en consonancia con la distinción que hicimos en el concepto. Esto es, analizar su papel desde el punto de vista tradicional (todo metadatos) y desde la perspectiva de su aplicación emergente para la identificación, descripción y control de la información electrónica:

PAPELES TRADICIONALES	PAPELES EMERGENTES
<ol style="list-style-type: none"> 1. Identificación y descripción de la información 2. Búsqueda y recuperación 3. Ubicación de los documentos 	<ol style="list-style-type: none"> 4. Autoría y propiedad intelectual 5. Formas de acceso 6. Actualización de la información 7. Preservación y conservación 8. Limitación del uso 9. Valoración del contenido 10. Visibilidad de la información 11. Accesibilidad de los contenidos

Tabla 7. Doble perspectiva del papel de los metadatos

Todas estas aplicaciones de los metadatos no son excluyentes, ni siquiera están ordenadas por orden de prioridad, pero sí demuestran su importancia creciente a tenor de la información electrónica y de la concepción de los documentos digitales como objetos de información. Con la tabla 7 tratamos de mostrar que el papel que desempeñan los metadatos en la gestión de la información digital es el mismo que han desempeñado siempre los catálogos y sistemas de descripción bibliográfica, la identificación, búsqueda y localización de la información, además de otras muchas funciones que demanda la idiosincrasia de la publicación en la Web. Podemos hablar pues, al igual que en el concepto, de continuidad en la aplicación de los metadatos. A pesar de ello, debemos constatar que la visión de los metadatos en un entorno informativo eminentemente digital, alberga una mayor complejidad y riqueza funcional.

En primer lugar, las funciones de los metadatos pueden enfocarse desde el punto de vista del sistema y desde el punto de vista del usuario. A nivel sistema, los metadatos pueden utilizarse para facilitar la interoperabilidad y la posibilidad de compartir datos entre las distintas herramientas de búsqueda, o incluso para facilitar una búsqueda híbrida e integral, tanto entre recursos de Internet, como en los materiales impresos previamente representados en MARC o en otro formato legible por máquina. A nivel usuario final, los metadatos deben facilitar detalles sobre qué información está accesible, dónde, cómo y en qué condiciones, se puede llegar a ella.

Dempsey y Heery⁷¹ plantean una reflexión sobre qué necesitan saber los distintos tipos de usuarios acerca de los recursos electrónicos heterogéneos a los que acceden en cada caso, que nos dará una primera aproximación a la aplicación de los metadatos en función de las necesidades del *cliente* de información Web. Estos autores señalan que el usuario necesita:

⁷¹ Lorcan Dempsey and Rachel Heery. Metadata: A Current View... *Op. cit.*, p. 148.

- En primer lugar, reconocer la existencia de recursos, es decir, recuperar o *descubrir*⁷² fuentes de información.
- Un usuario como investigador, precisa una información más amplia sobre un recurso de lo que puede ser una mera descripción, esto es, saber si le va a ser útil en su contexto profesional y/o de investigación. Para ello necesita conocer el origen, historia o responsabilidad intelectual de la fuente, su integridad y autenticidad, sus relaciones con otros recursos, las características específicas de un dominio particular en que se aplica tal información, etc.
- El usuario, visto como futuro proveedor de información en un proceso de retroalimentación constante en la producción de conocimiento, necesita saber también los derechos de propiedad intelectual vinculados a la información, y qué nivel de uso o reproducción puede hacer de él.
- Requiere saber también, si un recurso se ajusta a diferentes niveles de uso; si puede extraer el contenido con alguna herramienta o aplicación específica; conocer qué tipo de publicación es, esto es, si el documento es un manual, una monografía científica, una obra de divulgación, etc.
- Además, el usuario como consumidor de información, necesita saber bajo qué condiciones puede acceder a un determinado objeto de información, por ejemplo, si tiene que pagar por él y en qué términos y con qué cláusulas debe hacerlo.
- El usuario, como cliente o agente, necesita conocer las interfaces técnicas, los protocolos de acceso, los tipos o formatos de búsqueda, etc. que permite el sistema de información al que accede.

⁷² Vid. Glosario, *discovery*.

- Un usuario, como adulto, precisa también una valoración sobre los contenidos de un determinado servidor de información, por ejemplo, si incluye materiales que probablemente considere inadecuados para los menores.

Los esquemas de metadatos deben recoger todas estas necesidades de información del usuario que, básicamente podríamos resumir en dos: el acceso a la información y la utilidad de la información. A pesar del reconocimiento de estos requisitos y de que, a través de las diferentes definiciones que hemos apuntado, ya se trasluce la finalidad genérica de estos datos, así como un entorno difuso de usos diversos, la función principal de los metadatos, como seguiremos reafirmando a lo largo de este trabajo, es la recuperación de información. A pesar de ello, otra aplicación importante es la del control de los recursos digitales a través, tanto de metainformación para controlar el acceso, como de metadatos para calificar el contenido o para describir los derechos del documento. No obstante, podemos concretar de manera más sistemática, siguiendo a Lagonze y otros investigadores, que las principales aplicaciones de los metadatos, son⁷³:

- Catalogación (documentos, artículos y colecciones).
- Recuperación de información en Internet (*information discovery*).
- Comercio electrónico.
- Agentes de software inteligentes.
- Firmas digitales.
- Valoración y clasificación del contenido.

⁷³ Carl Lagonze, et al. Using Web Metadata: Part 2, the Resource Description Framework [presentación PPT]. *WWW7 Tutorial*. OCLC, 10 de abril de 1998. Disponible en: <http://purl.oclc.org/~emiller/talks/www7/tutorial/part2/sld005.htm> (consultado el 8 de julio de 1998), slide 5 y ss.

- Derechos de propiedad intelectual.
- Política de preservación y privacidad del contenido Web.

La mayoría de los autores de este campo reconocen estas aplicaciones de los metadatos, añadiendo otras o elidiendo alguna de ellas, según el ámbito en donde pongan el acento de sus investigaciones. Por ejemplo, Lupovici y Masanès⁷⁴, señalan dos grandes utilidades: la representación y la preservación de la información, ya que fundamentan su estudio en la necesidad de los metadatos para crear un sistema de depósito de las publicaciones electrónicas, semejante al del Depósito Legal en Bibliotecas Nacionales. Por otra parte, Taylor⁷⁵ incorpora otros usos que justifican el desarrollo de metadatos: el control administrativo y la gestión de la información. Comentaremos, no obstante las aplicaciones más importantes, prestando especial atención a su utilización en recuperación de información, fundamentalmente en bibliotecas digitales.

⁷⁴ Catherine Lupovici, Julien Masanès. *Metadata for Long-Term Preservation* [documento PDF]. NEDLIB Project. La Hogue: National Library of Netherlands, Koninklijke Bibliotheek, 29 de julio de 2000. Disponible en: <http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>, p. 10 y ss. (consultado el 24 de febrero de 2001). Este documento pertenece al interesante proyecto NEDLIB (*Networked European Deposit Library*) que tiene como objetivo preservar la documentación electrónica. En este sentido tratan de definir los metadatos mínimos necesarios para gestionar la conservación y/o preservación, de tal forma que se puedan manejar grandes cantidades de ítems en un entorno tecnológico. Para ello se basan en el Modelo de Referencia OAIS (*Open Archival Information System*), que si bien no es estrictamente un modelo de metadatos, funcionalmente actúa como tal, definiendo tipos diferentes de información conceptual.

Muir ha hecho una revisión de los proyectos en curso tendentes a arbitrar políticas y estrategias para preservar a través del tiempo, el patrimonio cultural e intelectual de las publicaciones electrónicas. Adrienne Muir. *Legal Deposit of Digital Publications: A Review or Research and Development Activity*. En: *Joint Conference on Digital Libraries (2001. Roanoke)*. New York: ACM, 2001, p. 165-173.

⁷⁵ Chris Taylor. *An Introduction to Metadata* [documento HTML]. Brisbane: University of Queensland, rev. 1 de abril de 1999. Disponible en: <http://www.library.uq.edu.au/iad/ctmeta4.html> (consultado el 2 de agosto de 2000).

El empleo de metadatos para optimizar la recuperación de información en Internet, a pesar de que Lagonze y sus colegas lo señalen en segundo lugar, es la aplicación más importante de los metadatos, reconocida por todos los autores y el punto fundamental y la hipótesis de partida de este trabajo. Coincidimos en que los metadatos aumentan la probabilidad de que los usuarios sean capaces de recuperar información apropiada para sus consultas y puedan evaluar su relevancia, utilidad y accesibilidad. Una descripción de los recursos Web bien estructurada y rigurosa gracias a la utilización de metadatos, puede aumentar la exactitud de la búsqueda y un orden de relevancia más preciso, de tal forma que posibilita una mayor automatización de las consultas y hace que no sea necesaria una validación manual de los resultados, supuestamente relevantes, que devuelven los sistemas de recuperación de información.

La utilidad de los metadatos para la recuperación está, como venimos precisando, reconocida en toda la bibliografía. Pero es importante destacar que no hemos encontrado ningún autor que haga referencia, salvo de soslayo o manera demagógica como parte del discurso de la Sociedad de la Información, a la utilidad de estos para la accesibilidad a los contenidos Web. Sin embargo las pautas para el desarrollo de contenidos Web accesibles (WCAG⁷⁶), recomendación emanada de la iniciativa WAI (*Web Accessibility Initiative*) del W3C, destacan que proporcionar metadatos para añadir información semántica a las páginas y sitios Web aumenta su accesibilidad. La inclusión de metadatos tiene un nivel de prioridad 2 de los tres niveles establecidos en dichas pautas para la verificación de la accesibilidad. Cuestiones como el tipo de información o las restricciones de acceso de la página,

⁷⁶ World Wide Web Consortium. *Web Content Accessibility Guidelines 1.0: W3C Recommendation 5-May-1999* [documento HTML]. Wendy Chisholm, Gregg Vanderheiden, Ian Jacobs, eds. W3C, 5 de mayo de 1999. Disponible en: <http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505> (consultado el 10 de enero de 2001). Para la creación de metadatos, las pautas del WAI remiten expresamente a las especificaciones del *Resource Description Framework* del W3C, cuyo modelo y sintaxis es recomendación desde ese mismo año. Vid. 5.2.3.

incluidas a través de metadatos en el código fuente de un documento Web, pueden ser de gran utilidad para las personas con discapacidades físicas, sensoriales o cognitivas, por ejemplo, y de forma más representativa, para los invidentes.

En cuanto la aplicación de metadatos para la catalogación de documentos y/o colecciones de DLOs, podría entenderse estrechamente relacionada con su uso para la recuperación de información; incluso, ya hemos señalado cómo algunas definiciones identifican la catalogación tradicional como una forma más de metadatos. Sin embargo, el hecho de utilizar los metadatos para la catalogación de recursos Web, no implica que éstos impliquen los elementos (áreas) tradicionales de encabezamiento, título, mención de responsabilidad, etc., los que se extraigan *a posteriori* por quien asigna la metainformación. Conllevan la descripción del contenido individual de los recursos en el momento de la creación del objeto Web (sea éste un ítem o conjunto de ítems de cualquier tipo o formato, que un ordenador puede usar o manipular como si se tratara de un objeto único). En el uso de los metadatos para catalogación o descripción de los recursos electrónicos, destacan las iniciativas como la llevada a cabo por el *Dublin Core* —que explicaremos más adelante— que tratan de remedar, de alguna manera, los campos o áreas de descripción que se recogen en las ISBD, a fin de describir, desde el momento de creación o momento cero de un recurso electrónico, todos los datos necesarios para su reconocimiento, descripción y recuperación posterior.

Otro de los ámbitos de aplicación de los metadatos, relacionado con la recuperación de información en la Red, que señalan Lagonze y otros autores⁷⁷, es su utilización por los agentes inteligentes. Los agentes *pueden considerarse como un software personal que asiste* en la búsqueda de información en la Red *con una*

⁷⁷ Lagonze, et al. *Op. cit.*, <http://purl.oclc.org/~emiller/talks/www7/tutorial/part2/sld010.htm>

*autoridad delegada de sus usuarios*⁷⁸. Su nombre se deriva de la facultad que se les presume de realizar acciones de una forma más o menos autónoma. En este contexto, los metadatos se utilizan para la representación, distribución e intercambio de conocimiento, así como para una comunicación efectiva, bien entre el usuario y el agente, bien entre el agente y el servicio de información asociado a él. Los metadatos aquí otorgan a los agentes posibilidad de entender el entorno en que se utilizan.

El resto de las aplicaciones que tienen los metadatos en el contexto de la información electrónica, están relacionadas fundamentalmente, con la conservación, la seguridad y fiabilidad de los recursos para una codificación de la información, sobre todo en el contexto del incipiente comercio electrónico⁷⁹, o con aspectos legales relativos al derecho a la privacidad (tanto de datos de carácter personal como de la información contenida en un sitio Web) y a los problemas de propiedad intelectual del conocimiento distribuido de esta forma. El estudio pormenorizado de todos estos usos no forma parte de los objetivos de esta investigación, pero haremos alusión a ellos ocasionalmente, reconociendo también su importancia. Asimismo también nos referiremos a su utilidad para caracterizar y juzgar los contenidos Web a través de una serie de etiquetas a las que se añaden valores numéricos que simbolizan una evaluación de los contenidos (violencia, pornografía, etc.⁸⁰).

⁷⁸ Fah-Chun Cheong. *Internet Agents: Spiders, Wanderers, Brokers and Bots*. Indianapolis: New Rider Publishing, 1996, p. 5. Cfr. Glosario, *searchbot*.

⁷⁹ Sobre este aspecto, Vid. Godfrey Rust. Metadata: The Right Approach. An Integrated Model for Descriptive and Rights Metadata in E-commerce [documento HTML]. *D-Lib Magazine*, July-August 1998, vol. 3, n° 7/8. Disponible en: <http://www.dlib.org/dlib/july98/rust/07rust.html> (consultado el 4 de agosto de 1998).

⁸⁰ Este es el uso específico que tiene el sistema de metadatos, elaborado por el W3C, PICS (*Platform for Internet Content Selection*) <<http://www.w3.org/PICS>>, cuya funcionalidad será absorbida poco a poco por RDF (*Resource Description Framework*). (Vid. 5.2.3). El formato PICS pertenece al tipo de metadatos diseñado inicialmente para ayudar a los padres y educadores en el control del acceso infantil a los contenidos de Internet, pero su etiquetado también facilita otros usos relacionados con la privacidad, seguridad, etc. de los recursos de la Red. Los metadatos PICS serán la base de la clasificación de contenidos Web que realizan instituciones como el ICRA (*Internet Content Rating*) [cont.]

Otra reflexión interesante en torno al para qué se aplican metadatos es la que nos proponen Grötschel y Lügger⁸¹ basándose en el principio de que la utilidad substancial de los metadatos es la de ser sustitutos de un ítem original. Estos autores identifican una serie de documentos o tipos de información digital (DLOs) donde es absolutamente necesaria la aplicación de metadatos. Según esto, podemos definir la aplicabilidad de la metainformación basándonos ahora, no en las necesidades de los usuarios sino en los requerimientos específicos de determinados tipos de información. Así, los metadatos se usarán con relación a:

- Registros con atributos de evidencia, dentro de los cuales se incluyen tesis, disertaciones, patentes y obras de arte.
- Documentos únicos o ítems protegidos: colecciones de museos, documentos o libros históricos, notas de conferencias, libros científicos, registros de audio y vídeo.
- Información relativa a condiciones comerciales como: gestión de documentos u órdenes en línea.
- Colecciones de datos de archivo o estadísticos, pertenecientes a autoridades gubernamentales, estadísticas o a administraciones locales.

En todas estas categorías documentales la información original no se distribuye libremente, ya que normalmente forma parte de colecciones privadas, de archivos o de depósitos de datos de una Intranet, por ejemplo, sin acceso libre; por ello es

Association) una organización internacional independiente, promovida por el RSAC (*Recreational Software Advisory Council*) que otorga poderes al público, especialmente a los padres, para realizar decisiones bien fundadas en cuanto a medios electrónicos gracias al etiquetado abierto y objetivo de contenidos para proteger a los niños contra las informaciones potencialmente nocivas, al mismo tiempo que salvaguarda la libertad de expresión en Internet.

⁸¹ Martin Grötschel, Joachim Lügger. *Op. cit.*, <http://elib.zib.de/ftp/pub/UserHome/Luegger/Dresden/Metadata.htm>

interesante contar con un registro de metainformación que pueda distribuirse libremente por Internet, en lugar del documento original. Esta es pues otra aplicación reconocida para los metadatos, ser sustituto o anticipo de la información original que presenta condiciones de autenticidad o confidencialidad protegidas.

Algo semejante ocurre en el ámbito digital de los negocios, más comúnmente conocido como comercio electrónico. Para realizar legalmente transacciones comerciales de forma electrónica es necesario identificar tres piezas claves de metainformación: el usuario, el objeto de la compra y el uso que se va a hacer de ella. Sin profundizar en esta utilidad de los metadatos, es oportuno evocar el proyecto INDECS <<http://www.indecs.org>> que establece un conjunto de metadatos que se denomina acuerdo (*agreement*), donde se identifica información sobre las partes y los papeles que desempeñan (el objeto al que se refiere el contrato, la acción, los parámetros que limitan la acción, etc.). La aplicación de los metadatos al comercio electrónico es otra práctica habitual que denota una vez más su utilidad en el contexto de la información electrónica y, en este caso, de las transacciones comerciales que surgen en su seno.

Otra forma de hacer útiles los datos a través de metainformación —como veíamos al plantear el enfoque de la gestión de datos en Burnett, Bor Ng y Park⁸²— es analizarlos. Una aplicación interesante en este sentido se puede traslucir de los trabajos bibliométricos que se están realizando en y sobre la Web. Por ejemplo, Almind e Ingwersen⁸³ afirman que la WWW es una red de citas y aplican las

⁸² Katheleen Burnett, Kwong Bor Ng, Soyeon Park. *Op. cit.*, p. 1210, 1212.

⁸³ Tomas C. Almind and Peter Ingwersen. Informetric Analysis on the World Wide Web: Methodological Approaches to «Webometrics». *Journal of Documentation*, 1997, vol. 53, nº 4, p. 404-426. Citado en: Chun Wei Choo, Brian Detlor and Don Turnbull. *Web Work: Information Seeking and Knowledge Work on the World Wide Web*. Dordrecht, Boston, London: Kluwer Academic Publishers, 2000, p. 144.

Desde hace algún tiempo son muchos los autores que están intentando hacer análisis métricos de la WWW a través de la utilización de sistemas de indización y búsqueda de información que, como
[cont.]

técnicas informétricas a las páginas Web, comparando algunas de las etiquetas HTML con la información sobre citación establecida en las bases de datos del ISI (*Institute for Scientific Information*). Los estudios "webométricos" que se están haciendo hoy en día, basándose en las posibilidades de búsqueda por dominios, URLs, etc. que permiten algunos sistemas de recuperación de información de Internet, son aún bastante imperfectos o poco fiables. Con esta afirmación se puede advertir una utilidad implícita de los metadatos, la posibilidad de caracterizar el contenido de las páginas Web en aras a facilitar el análisis métrico de ese nuevo soporte de información científica y técnica. El hecho de utilizar metadatos aumentará el grado de homologación de la información, que es uno de los principales problemas que tienen los buscadores como Altavista al no existir una estandarización de metadatos en las cabeceras de las páginas, que permitan a estos sistemas indizar sólo aquellos campos (autor, título, palabras clave, resumen, etc.) susceptibles de interrogación posterior para este tipo de análisis.

Kashyap y Sheth⁸⁴ por su parte, destacan la utilidad de los metadatos en el *brokering*, es decir, en el proceso de mediación entre los usuarios y los proveedores de información, como una solución clave para afrontar el problema de la sobrecarga informativa, salvar el problema de la heterogeneidad y hacerla accesible, teniendo

Altavista, permiten realizar búsquedas cualificadas por campos. (V. gr. Josep Manuel Rodríguez i Gairín. Valoración del impacto de la información en Internet: Altavista, el «Citation Index» de la Red. *Revista Española de Documentación Científica*, 1997, vol. 20, nº 2, p. 175-181; o Ray R. Larson. *Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace*. En: *ASIS Annual Meeting (59. 1996. Baltimore, Maryland)*. Medford: Information Today, 1996, p. 71-78. Estos estudios tentativos indican la necesidad de medir el factor de impacto, realizar análisis de cocitación, etc. y demás indicadores métricos para la Web como soporte de información científica. No obstante, hasta ahora, los resultados que producen los buscadores (tipo Altavista o WebCrawler, sistemas que se analizan, respectivamente, en los artículos que citamos) fruto, por ejemplo, de una consulta temática o sobre el autor, pueden ser tan variados y tan peregrinos como la concatenación de caracteres introducida en la pregunta lo permita.

⁸⁴ Virpul Kashyap, Amit Sheth. *Information Brokering Across Heterogeneous Digital Data: A Metadata-based Approach*. Boston, Dordrecht, London: Kluwer Academic Publishers, 2000. Para una explicación más exhaustiva del término *information brokering*, Vid. Glosario.

como resultado una reducción del volumen y de los datos que deben tratarse. En ese proceso, los metadatos tienen una utilidad básica que proporciona una nueva arquitectura de la información. El *brokering* incide en la recuperación e implica la creación y la gestión de sistemas de información eficaces que permitan entender y explotar el significado y el uso de la información y facilitar el intercambio de ésta, al mismo tiempo que atiende a las necesidades de los usuarios. La clave del *brokering* es capturar el contenido de una forma independiente a su representación, por ello, las descripciones de metadatos serán fundamentales para extraer dicho contenido.

Además Kashyap y Sheth⁸⁵ plantean, de una forma más sistemática y concreta, una doble perspectiva en la gestión de metadatos: por un lado, una visión basada en los escenarios de aplicación que envuelven a los datos multimedia, que nos servirá para profundizar ahora en torno a los usos de los metadatos; y una perspectiva basada en el contenido de la información, que nos servirá para tipificarlos en el apartado siguiente (3.5).

En todos los escenarios de aplicación que advierten estos autores para los metadatos está implícita la utilidad fundamental de la recuperación, sin embargo señalan una casuística de uso en sistemas de información concretos en el contexto del *brokering* de información. Estos escenarios son:

- *Navegación, browsing y recuperación de colecciones de imágenes*: Se alude al uso de los metadatos como soporte a una navegación y recuperación semántica de la imagen. En aquellos sistemas de información que mantienen amplias colecciones de imágenes se requiere un conocimiento sobre el contenido espacial de las mismas y de su comportamiento.
- De forma semejante, advierten el uso de metadatos para recuperar imágenes en movimiento extraídas de un *video*, cuya búsqueda debe basarse también en el

⁸⁵ Virpul Kashyap, Amit Sheth. *Ibid.*, p. 18 y ss.

contenido semántico, captando el contenido espacial y también el aspecto temporal.

- Igualmente, señalan el uso de los metadatos asociados a los registros de *audio*, por ejemplo en sistemas de información de radio, de tal forma que se capture, en forma textual, la esencia del contenido sonoro.
- La *gestión de documentos estructurados*. En este sentido, señalan la utilidad de los metadatos en el procesamiento de documentos estructurados, advirtiendo la importancia que están tomando formatos como SGML y XML en el cambio de paradigma de la publicación, que introducen estructura y metadatos. La recuperación eficiente se consigue a través de la explotación de esa estructura y de la utilización de los metadatos para la indización.
- *Los sistemas de información geográfica y medioambiental*. Este tipo de sistemas tienen una amplia variedad de usuarios con necesidades de información específicas. El requisito clave en este tipo de sistemas de información es la integración, que será posible a través de información descriptiva dirigida tanto al sistema como al usuario final. Esto implica la descripción de metadatos, por un lado y la cohesión de vocabularios utilizados por los distintos sistemas de información para interpretar las descripciones, por otro.
- *Acceso a medios diversos*. En este contexto se refieren a la necesidad de acceder a distintos tipos de información (texto, imagen fija o en movimiento, etc.) donde será preciso contar con sentencias de búsqueda formuladas independientemente del tipo de medio, por ello, es necesario utilizar descripciones de metadatos para detallar, de manera uniforme, la información de los distintos medios.

Hemos dejado para el último lugar, dentro de los escenarios de aplicación que señalan estos autores, las *bibliotecas digitales*, ya que es el ámbito fundamental en el que nos centraremos en esta tesis doctoral. Kashyap y Sheth, reconocen que las bibliotecas digitales ofrecen una amplia gama de servicios y colecciones de documentos digitales, y destacan —como nosotros en este trabajo— que constituyen

un área de aplicación llena de retos para el desarrollo e implementación de esquemas de metadatos. La clave para integrar las diferentes colecciones en una biblioteca digital es el uso de metadatos que representen las características inherentes y contextuales de una colección digital.

Katrina Hodgson reconoce también el amplio espectro de aplicación de los metadatos para proporcionar funcionalidad en la nueva infraestructura de la información. De forma genérica, siguiendo a esta autora⁸⁶, se nos ocurre recapitular los usos de los metadatos, aludiendo nuevamente a la doble aproximación al concepto de Burnett, Bor Ng y Park:

- Desde el enfoque del control bibliográfico, los metadatos pueden usarse para resumir, buscar, considerar y evaluar, seleccionar, recuperar y especificar el uso e interpretación de los datos a los que se refieren, así como para describir la historia y origen de los mismos.
- Desde el punto de vista de la gestión de los datos, sirven para administrar y gestionar un DLO dentro de un servidor, para definir las relaciones entre los distintos objetos de datos y para especificar la estructura y los componentes lógicos de los objetos de información y el modo de acceso a ellos.

Recogiendo todo lo dicho hasta aquí, y basándonos en la representación de Granger⁸⁷, podemos interpretar la aplicación de los metadatos desde una triple perspectiva, que nos servirá asimismo para tipificarlos a continuación: desde la

⁸⁶ Katrina Hodgson. *Op. cit.*, <http://www.slis.ualberta.ca/538/khodgson/metadata.htm>. En este interesante trabajo, Hodgson utiliza las mismas fuentes que nosotros (Lagonze, Ianella y Waugh, etc.) y propone un listado de aplicaciones de los metadatos en el contexto de la información electrónica que resume, de forma contundente y concisa, los usos de los mismos. La división de estas aplicaciones, retomando la doble aproximación de Burnett, Bor Ng y Park que citamos en el apartado 3.2. dedicado al concepto, es nuestra.

⁸⁷ Stewart Granger. Metadata and Digital Preservation: A Plea for Cross-interest Collaboration. *Vine (Theme issue: Metadata. Part 2)*, 1999, n° 117, p. 26.

visión del dominio científico que los utilice, desde el punto de vista de su uso genérico y desde la perspectiva del papel que desempeñan en la gestión de la información o en la creación de datos (Fig. 1). Esta triple visión nos permite asimilar el amplio espectro de aplicación de los metadatos, que hemos reflejado a lo largo de este apartado, a tres secciones conceptuales: describir, recuperar y comunicar/compartir información.

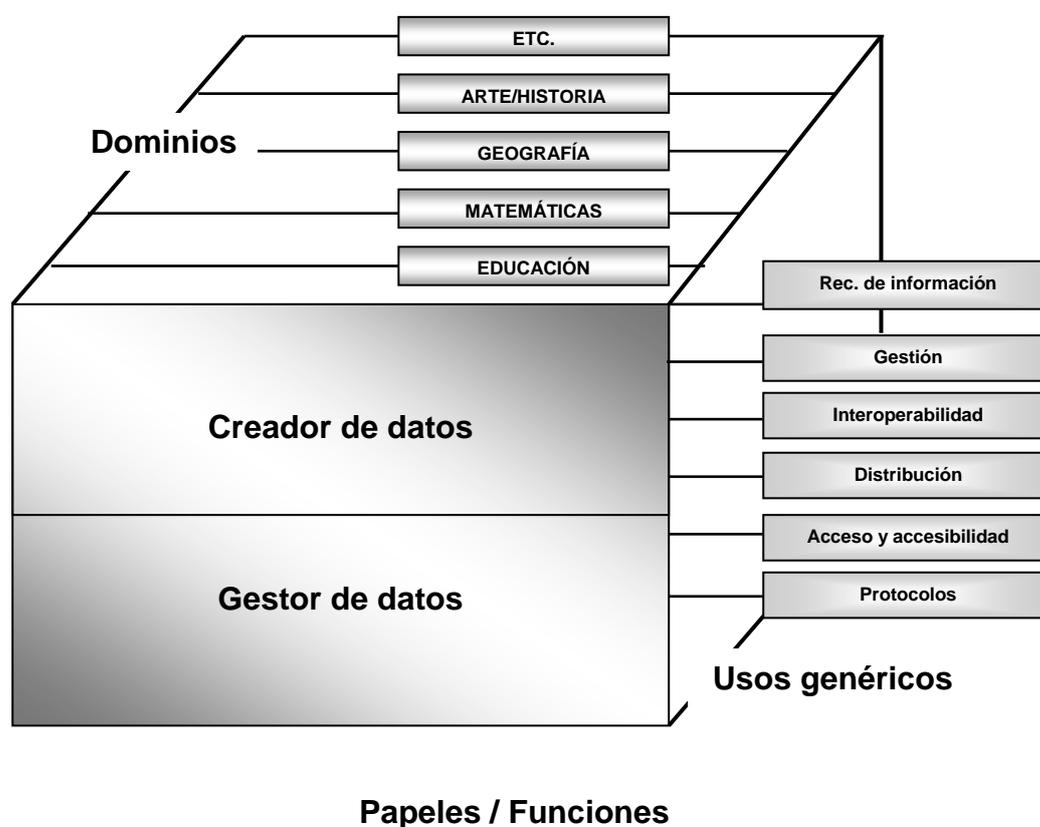


Fig. 1. Triple visión de la aplicación de los metadatos

3.5. Tipos de metadatos

Con lo que hemos dicho hasta el momento, parece obvio que existen distintos tipos de metadatos en la medida que existen diversos tipos de usuarios, con necesidades distintas y diversos tipos de información, con finalidades también distintas. Según esto, trataremos ahora de tipificar o clasificar los metadatos desde

diferentes puntos de vista, algunos ya vislumbrados en la exposición anterior y otros nuevos.

Con la amplitud que hemos dado al concepto hasta ahora, son múltiples las clasificaciones de los metadatos que podríamos hacer, basándonos en su origen, forma, funcionalidad, estadística de uso, nivel de estructuración de los datos, etc. Sin perjuicio de usar también estos criterios, una de las clasificaciones más sencillas que podemos hacer, en correlación con el apartado anterior, es una división de los metadatos según el ámbito de aplicación en que se utilizan. De este modo, y de forma semejante a la perspectiva del escenario de aplicación de Kashyap y Sheth, que tratamos antes o, dicho de otra forma, teniendo en cuenta la dimensión de la metainformación según el dominio de aplicación (Fig. 1), podemos decir que existen metadatos:

- Para describir recursos (objetos) de información Web asociados al concepto de biblioteca digital. Este es el caso, por ejemplo del modelo propuesto por el Dublin Core (DCMI).
- Metadatos que definen los elementos de los registros catalográficos en bibliotecas, como es el caso del formato MARC⁸⁸ (*Machine Readable Cataloging*).

⁸⁸ El formato MARC puede entenderse, en sentido amplio, como un modelo más de metadatos. Es preciso comentar que a pesar de ser un formato antiguo, creado inicialmente para compartir la tediosa tarea de creación de registros en formato electrónico, existe una tendencia —concretamente en el contexto americano del USMARC— protagonizada por el MARBI (*Machine-Readable Bibliographic Information*) para incluir, en dicho formato, la descripción de recursos electrónicos a través del campo 856. La finalidad de los proyectos de catalogación de recursos electrónicos en MARC es completar las colecciones existentes actualmente en las bibliotecas, así como fomentar la recuperación de información por el usuario final y mejorar los requerimientos de los sistemas actuales de bibliotecas, de tal forma que se adapten a los cambios y puedan manipular los registros. A lo largo de este trabajo tendremos oportunidad de tratar diversos temas relacionados con la adecuación del formato MARC para describir DLOs (*Vid.* 4.1. y 6.1.).

- Metadatos asociados a la información de museos, y por tanto a una información eminentemente iconográfica, por ejemplo, el modelo del CIMI (*Consortium for the Interchange Museum Information*).
- Para recursos geográficos o geospaciales, como por ejemplo el *Content Standard for Digital Geospatial Metadata (CSDGM)*⁸⁹ del FGDC, (*Federal Data Geographic Committee*), que constituye una norma para la descripción de datos geospaciales. Dentro de este mismo campo de aplicación (información geoespacial), también se encuentra el DIF (*Directory Interchange Format*) del *Global Change Master Directory* de la NASA⁹⁰.
- Para describir recursos de información electrónica gubernamental, como en el caso del formato de metadatos GILS (*Government Information Locator Service*).
- Para la descripción archivística, como alternativa a la descripción tradicional realizada en los archivos. En el ámbito archivístico se ha reconocido también la capacidad de los metadatos para proporcionar información descriptiva en los procesos de creación de registros electrónicos, que puede llegar a obviar, o a reducir de manera significativa, la *catalogación* archivística tradicional. En este campo, podemos hablar del EAD (*Encoded Archival Description*) como modelo de metadatos basado en SGML/XML que representa una forma estructurada de crear ayudas para la búsqueda digital en colecciones constituidas por materiales de archivo.

⁸⁹ Para más información, Vid. Federal Geographic Data Committee. *Content Standard for Digital Geospatial Metadata (CSDGM)* [documento HTML]. Reston: FGDC, USGS, rev. 11 de diciembre de 2000. Disponible en: <http://www.fgdc.gov/metadata/contstan.html> (consultado el 10 de febrero de 2001).

⁹⁰ Sobre este estándar de metadatos para la información geoespacial del sistema de datos de la NASA, Vid. *Directory Interchange Format (DIF) Writer's Guide, Version 7, 1999* [documento HTML Global Change Master Directory. National Aeronautics and Space Administration, May 1999. Disponible en: <http://gcmd.gsfc.nasa.gov/md/difguide/difman.html> (consultado el 10 de febrero de 2001).

- Existen también sistemas de metadatos específicos en el área de desarrollo de dispositivos inteligentes para el almacenamiento de datos, tal es el caso de MEMRI⁹¹ (*Media Error Monitoring and Reporting Information*) destinados a proporcionar datos sobre la integridad de la información almacenada en este tipo de soportes inteligentes.

Otra clasificación posible, teniendo en cuenta ahora las distintas fases por las que pasa un documento electrónico entendido como objeto, es la división según la utilidad y/o función que prestan en esas diferentes etapas de lo que antes denominamos *ciclo vital de un objeto digital*. Así podríamos establecer cinco tipos de metadatos que se corresponden más o menos a las etapas del ciclo de un DLO señaladas por Guilliland-Swetland:

1. Metadatos administrativos: utilizados para la gestión y administración de los recursos digitales en red en el momento de su creación o introducción en el sistema. Este tipo de metadatos es esencial para comprobar el nivel de mantenimiento de los recursos Web; por ejemplo, la ubicación de la información, la institución en cuyo servidor se alberga, o el seguimiento y control de las distintas versiones.
2. Metadatos descriptivos: para la representación o identificación de los recursos de información en la fase de organización de los objetos digitales (por ejemplo, registros catalográficos, índices especializados, ayudas de búsqueda o URNs⁹²).

⁹¹ Fernando L. Podio, Willian Vollrath, Ben Kobler. Media Error Monitoring and Reporting Information (MEMRI). Metadata for Intelligent Digital Data Storage Devices [documento HTML]. En: *IEEE Metadata Conference (2. 1997. Maryland)*. IEEE Computer Society, 1997. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/fpodio/fpodio.htm> (consultado el 18 de junio de 1998).

⁹² URN= *Uniform Resource Name/Number* (Nombre/Número Uniforme de Recursos). Es la identificación única e independiente de cada archivo accesible en Internet. Los URNs tratan de mantener accesibles siempre los ficheros Web que identifican, independientemente de que se produzcan cambios en su servidor o ruta de acceso (*path*) específica.

3. Metadatos para la conservación⁹³: destinados a gestionar la preservación de las fuentes de información (por ejemplo, documentación sobre la renovación y actualización de los datos o la migración del contenido a otro sitio), en aras a resolver los principales problemas de la información digital, esto es, la fragilidad y el carácter efímero de los DLOs y el problema de la obsolescencia de los medios informáticos (hardware y software).
4. Metadatos Técnicos: creados por, o generados para, un sistema automatizado, relativos al funcionamiento de dicho sistema como, por ejemplo, información sobre el hardware o software necesarios para acceder a un determinado servicio de información electrónica, datos sobre los tiempos de respuesta, etc.

⁹³ Los proyectos más interesantes en este sentido son:

- NEDLIB <<http://www.kb.nl/coop/nedlib>> del que ya hemos hablado. *Cfr. suprà*, Nota 74.
- CEDARS <<http://www.leeds.ac.uk/cedars>>. Se trata de un proyecto del Consorcio de Bibliotecas Universitarias del Reino Unido (CURL) que trata de crear un prototipo de archivo digital que incluya una muestra de materiales representativos de colecciones digitales sobre los que se aplicará una DTD de XML con un conjunto básico de elementos de metainformación que aseguren la preservación de dichos objetos digitales.
- InterPARES <<http://is.gseis.ucla.edu/us-interpares>>. Proyecto norteamericano, originado por la *National Historical Publications and Records Commission*, para identificar y modelar la forma, función y estructura de los sistemas digitales, su contenido y los metadatos críticos para cubrir estas necesidades. Se centra sobre todo en la información creada para la conservación y preservación permanente de documentos electrónicos en distintos contextos organizacionales y sociales.
- PANDORA <<http://pandora.nla.gov.au>> de la Biblioteca Nacional de Australia, que diseña un modelo de datos lógico basado en el modelo entidad-relación, que ayude a identificar las entidades particulares y los metadatos asociados en su servicio de información prototipo.

Para una reflexión general sobre los metadatos de conservación, *Vid. Stewart Granger. Op. cit.*, p. 24-29. En este documento se habla, entre otras cosas, de algunas estrategias para preservar la información digital, tales como la actualización, migración, emulación y preservación tecnológica, y del papel que desempeñan los metadatos en ellas. Para ampliar la información de este tipo de metadatos, se recomienda la consulta del servicio de información temática que ha desarrollado la Biblioteca Nacional de Australia, PADI (*Preserving Access to Digital Information* <<http://www.nla.gov.au/padi>>) y más concretamente, la sección dedicada exclusivamente a los metadatos de preservación <<http://www.nla.gov.au/padi/topics/32.html>>.

5. Y por último, metadatos de uso, que se refieren a los metadatos, creados generalmente de forma automática, relativos al nivel de utilización y al tipo de usuarios de un determinado servicio de información electrónica.

Los cuatro primeros tipos de metadatos de esta clasificación están estrechamente relacionados con la identificación de las necesidades de información que tienen los distintos usuarios a la hora de enfrentarse a un sistema de información y con la descripción de las características principales de un DLO; el último tipo, sin embargo, hace referencia a la información de uso que controla el ordenador. Esta forma de dividir los metadatos en dos grandes grupos o categorías, aunque con variaciones y contradicciones como veremos, es la más común y recurrente a la hora de tipificar los metadatos.

Por ejemplo Prothman⁹⁴ hace eco de esa división biaxial y distingue: a) por un lado los metadatos que describen la entidad de datos o el objeto de información en sí mismo, y por otra, b) aquellos que describen su almacenamiento y uso. No obstante señala cinco categorías más específicas:

1. *Metadatos de acceso*: que son los que permiten la interrogación, navegación y recuperación de información; describen, entre otras cosas, cómo están los datos estructurados lógicamente.
2. *Metadatos semánticos*: aquellos que sirven para dotar a la información almacenada de un significado o propósito específico.
3. *Metadatos de calidad*: aquella metainformación recogida para permitir un análisis cualitativo de los datos.
4. *Metadatos de transferencia*. Este tipo correspondería en la clasificación anterior, basada en el ciclo de vida de los documentos de Guilliland-Swetland, a los

⁹⁴ Bill Prothman. *Op. cit.*, <http://www.ieee.org/membership/students/potentials/febmar2000/pdf/pages20.pdf>, p. 20. Nótese como también en la clasificación de Prothman se puede ver claramente implícita la aplicación de los metadatos que resumíamos en el apartado anterior.

metadatos técnicos, e igual que ellos, hace referencia a cómo los datos pueden transferirse entre aplicaciones.

5. *Metadatos de almacenamiento*: aquéllos que revelan cómo y dónde se almacenan los datos dentro de un sistema.

Burnett, Bor Ng y Park⁹⁵ distinguen tácitamente dos tipos de metadatos al comparar los atributos de distintos esquemas: a) *metadatos intrínsecos*, donde incluyen atributos como: materia, título, autor, editor, lugar de publicación, otro agente, fecha, tipo de objeto de información, forma del identificador (URN, ISBN), relación, fuente, idioma, cobertura, resumen, versión, notas, firma, clasificación, nivel de seguridad y descriptores; y b) *metadatos extrínsecos*, a los que pertenecen informaciones como: requerimientos del sistema, modo de acceso, accesibilidad, coste, control, extensión o tamaño del documento, descripción codificada y descripción de la revisión. En sentido laxo, los metadatos extrínsecos corresponderían a los datos que describen la entidad o el objeto de información en sí mismo, en la división genérica hecha por Prothman, y los extrínsecos, a los metadatos que describen el almacenamiento y uso.

De forma análoga, Hill y otros investigadores de la ADL⁹⁶ también distinguen dos tipos fundamentales de lo que ellos denominan *colecciones de metadatos* aplicadas a dicha biblioteca digital real. Diferencian así: a) *metadatos inherentes*, que comprenden información derivada del análisis informático de los contenidos de una

⁹⁵ Estos autores establecen un cuadro comparativo de atributos en seis esquemas de metadatos, concretamente: DC, URC, SH, USMARC, IAFA y TEI. No obstante y desde nuestro punto de vista esta información es aplicable a cualquier tipo de formato, máxime su división en atributos intrínsecos y extrínsecos. Kathleen Burnett, Kwong Bor Ng, Soyeon Park. *Op. cit.*, p. 1214.

⁹⁶ Linda Hill y sus colegas tratan los metadatos desde una perspectiva de aplicación concreta a una biblioteca digital (*Alexandria Digital Library*) y manejan el concepto de colección digital, para cuya gestión es indispensable las colecciones de metadatos. Sobre esta visión de la biblioteca digital profundizaremos en el capítulo 8. Linda L. Hill, et al. Collection Metadata Solutions for Digital Library Applications. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, nº 13, p. 1170.

colección digital, tales como: cobertura temporal, tipos de ítems, el número de cada uno de ellos, etc.; y b) *metadatos contextuales*, la información suministrada por el gestor de la colección, es decir, información que no puede extraerse directamente de los contenidos de la colección, por ejemplo: título, responsabilidad, fecha de creación, alcance y propósito, términos y condiciones de uso de la colección, etc. Esta categorización surgida de la *Alexandria Digital Library*, parece contradictoria con la anterior, pues los metadatos que Burnett, Bor Ng y Park denominan intrínsecos, corresponden, *grosso modo*, a los metadatos contextuales en el planteamiento de la ADL, y los metadatos inherentes en esta clasificación, podrían asimilarse a los metadatos extrínsecos en la tipología anterior. Estas disensiones en torno a los tipos de metadatos ponen de manifiesto algunas de las hipótesis que manejamos en esta investigación:

- Por una parte, corroboran que los metadatos aún están en fase de construcción técnica y que no existe un consenso ni en el concepto, ni en la tipificación. Simplemente se están sentando las bases de un planteamiento válido para optimizar la recuperación de información, pero la universalidad o estandarización global del planeamiento dista mucho de ser una realidad.
- Y por otra parte, también se ratifica que la aplicabilidad de los metadatos, en la concepción de los sistemas de información electrónica actuales, es factible en entornos finitos de información, ya sea el límite lógico (como el nivel de concreción informativa que supone una Intranet) o temático, como una biblioteca digital especializada, como es el caso de la ADL.

Darlene Fichter⁹⁷ también habla de dos tipos de metadatos, aunque circunscribe su categorización al entorno finito de las Intranets corporativas; no obstante, la tipología de Fichter es una de las más coherentes desde nuestro punto de vista y es

⁹⁷ Darlene Fichter. Administrative and Factual Metadata for Intranets: Issues and Options. *Online*, 1999, vol. 3, nº 6, p. 88.

perfectamente extrapolable al entorno multidimensional de Internet y de las bibliotecas digitales. Los dos grandes tipos de metadatos que distingue esta autora son:

- a) Por un lado, los que denomina *metadatos administrativos o factuales*, es decir la información bibliográfica y la de gestión que hace referencia a las circunstancias del documento, tales como la fecha de creación, la de modificación, el idioma del documento, el autor, el tamaño, etc.
- b) Y por otro, *metadatos estrictamente descriptivos*, que se centran en la descripción o creación de un sustituto para un objeto de información. Este tipo de metadatos está estrechamente vinculado al concepto tradicional de indización que lleva aparejado aspectos de control y normalización del contenido.

Si hacemos una reflexión elemental, partiendo del paradigma tradicional del análisis documental en bibliotecas, podríamos decir que los metadatos factuales son a la información electrónica lo que la catalogación o descripción bibliográfica a los materiales "tradicionales" en bibliotecas; y siguiendo la correlación, los metadatos descriptivos corresponderían al análisis de contenido. Lo realmente original e interesante de esta tipología es este último tipo de metadatos. Los metadatos, denominados aquí factuales, no son más que la metainformación contextual e inherente, en el planteamiento de la ADL explicado por Hill y sus colegas, o los metadatos intrínsecos y extrínsecos en la tipología de Burnett, Bor Ng y Park. Sin embargo los metadatos descriptivos aluden a un nuevo enfoque basado en el contenido.

En esta misma línea, una clasificación importante es la que se desprende también del trabajo de Kashyap y Sheth⁹⁸ al analizar los metadatos desde la

⁹⁸ Virpul Kashyap, Amit Sheth. *Information Brokering... Op. cit.*, p. 19-22. A pesar de que aquí tomamos como referencia la clasificación del último libro de Kashyap y Sheth (2000), estos mismos autores sostienen, aunque con algunas variaciones, la misma tipología en otras publicaciones, p. ej. [cont.]

perspectiva del contenido. Se basan fundamentalmente, en la cantidad de contenido informativo que engloban, e identifican, al mismo tiempo, el papel clave que desempeñan los metadatos en la interoperabilidad semántica. Así pues, distinguen —siguiendo la tendencia a la división biaxial que detectamos también en las otras clasificaciones— dos tipos fundamentales de metadatos:

- a) *Metadatos independientes del contenido*, que vienen a ser los factuales que señalábamos antes, es decir, los metadatos que recogen aquella información que no depende del contenido del documento, como la localización, la fecha de creación, modificación, etc. Kashyap y Sheth, al igual que expresaba Fichter para los metadatos factuales, reconocen la utilidad de estos datos referenciales para recuperar información y para clasificarla.
- b) *Metadatos dependientes del contenido*. Son, como cabe esperar, los metadatos que dependen del contenido asociado al objeto de información que describen. Este tipo de metadatos recoge normalmente información representacional y estructural y facilitan la interoperabilidad. Se subdivide en:

Virpul Kashyap, Amit Sheth. *Semantic Heterogeneity in Global Information Systems: The Role of Metadata, Context and Ontologies* [documento PS]. Georgia: University of Georgia. Department on Computer Sciences. Large Scale Distributed Information Systems, 12 de junio de 1996. Disponible en: <http://lsdis.cs.uga.edu/lib/download/KS97.ps> (consultado el 12 de febrero de 2001).

Además esa misma categorización de los metadatos parte de la literatura consagrada de las bases de datos multimedia, y ya había sido perfilada en un trabajo de estos mismos autores junto con Kshitij Shah en 1996. En aquella clasificación distinguían taxativamente tres tipos de metainformación: *metadatos independientes del contenido*, *metadatos dependientes del contenido*, y *metadatos descriptivos del contenido*. Estos últimos (descriptivos del contenido), en la clasificación que recogemos aquí, será un subtipo de los metadatos dependientes del contenido. V. Kashyap, K. Shah, and A. Sheth. Metadata for Building the Multimedia Patch Quilt. En: *Multimedia Database Systems: Issues and Research Firections*. V. S. Subrahmanian and Sushil Jajodia, eds., Srpinger, 1996, p. 297-319. Citado en: François Role. *Panorama des travaux en cours dans le domaine des métadonnées: Rapport de recherche n° 3628*. [documento PDF]. Paris: INRIA, février 1999. Disponible en: <ftp://ftp.inria.fr/INRIA/publication/publi-pdf/RR/RR-3628.pdf>, p. 7 (consultado el 12 de julio de 1999).

- *Metadatos basados en el contenido directo*, esto es, aquellos que dependen directamente del contenido del objeto de información. Por ejemplo, los índices de un documento en texto completo o el color y la forma en una imagen digital.
- *Metadatos descriptivos del contenido*. Este tipo de metadatos describe la información de un documento o DLO sin utilizar expresamente su contenido, como por ejemplo, las anotaciones textuales (descriptores, identificadores, etc.) que especifican el contenido de una imagen. Este tipo de metadatos, a su vez, se pueden entender en un doble sentido⁹⁹:
 - *Metadatos independientes del dominio*, que recogen la información presente en el documento sin tener en cuenta dominio de aplicación¹⁰⁰ y que tienen una naturaleza eminentemente estructural. Normalmente estos metadatos constituye la base de la indización de las colecciones de documentos (DLOs) y permiten una recuperación más rápida. Ejemplos de este tipo de metadatos podrían ser: las diferentes DTDs (*Document Type Definition*) de SGML/XML que definen una estructuración de materias de contenido que permite optimizar la eficacia en la recuperación de información.
 - *Metadatos de un dominio específico*. Son aquellos que adoptan una forma especial dependiendo del campo de aplicación o de la materia de la

⁹⁹ La subdivisión de este tipo de metadatos que, siguiendo a Kashyap y Sheth, detallamos a continuación tiene en esencia la misma fundamentación de la división que plantearemos nosotros en el capítulo 5 para describir los modelos o formatos generales de metainformación. En nuestro planteamiento señalaremos, por un lado, formatos de propósito general (*independientes del dominio*) y formatos con propósitos específicos (*dependientes de un dominio específico*, en la denominación de estos autores). Vid. 5.2.1 y 5.2.2., respectivamente.

¹⁰⁰ Recordemos aquí los escenarios de aplicación que contemplan, entre otros, los propios Kashyap y Sheth (*Cfr. suprà.* 3.4.), así como la primera clasificación de los metadatos que hemos hecho en este apartado basándonos en el dominio temático de aplicación de los mismos.

información que describen. Este es el caso de metadatos como la cobertura geográfica —atributo `land-cover`— en sistemas de información geográfica, como la ADL; o metainformación relativa a la población, en el dominio de la información estadística. En este tipo de metadatos es fundamental el vocabulario ya que es específico de cada ámbito o dominio informativo, y la forma de consignar la información de una forma significativa es también particular de cada campo temático. Los metadatos específicos de un dominio pueden construirse a partir de ontologías específicas, o de términos que describen la información en bibliotecas conceptuales. De esta forma, las ontologías se pueden entender como metadatos, cuyas relaciones terminológicas serán indispensables para conseguir la interoperabilidad semántica en el nivel del vocabulario¹⁰¹.

Una ratificación de esta misma clasificación, la encontramos en la reciente obra *Professional XML Meta Data*¹⁰² donde diversos autores reconocen, desde una perspectiva más técnica, también dos tipos de metadatos según su propósito: a) metadatos basados en el recurso (*resource based meta data*) cuya función es la catalogación y la identificación y están constituidos por propiedades de los recursos

¹⁰¹ Sobre la interoperabilidad, *Vid.* 8.3.2. y Glosario, *interoperability, crosswalks/crossroads*. Mencionaremos previamente también este aspecto al hablar de la normalización del vocabulario con relación a los metadatos (*Vid.* 6.4.). No obstante, la creación de ontologías es uno de los mayores retos de la investigación actual de la Inteligencia Artificial. La ontología es la parte de la metafísica, que trata del ser en general y de sus propiedades trascendentales, en el contexto de la organización y representación del conocimiento es, como dice Gruber, *una especificación de una conceptualización*, una descripción de los conceptos y de las relaciones que existen entre ellos en un dominio particular. Tom Gruber. *What is An Ontology* [documento HTML]. Stanford: Stanford University, Knowledge Systems Laboratory, rev. 12 de septiembre de 1997. Disponible en: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (consultado el 11 de febrero de 2001). Una ontología está constituida por reglas para crear un diccionario de descripciones de conceptos, por ello las consideramos (junto a otras herramientas terminológicas que relacionan conceptos, como los tesauros) metadatos descriptivos del contenido o basados en materias.

¹⁰² Kal Ahmed, et al. *Professional XML Meta Data*. Birmingham: Wrox Press, 2001, p. 13-14.

de información; y b) metadatos basados en la materia (*subject based meta data*), formados por datos que representan el contenido y sus relaciones, así como recursos de información específicos que pertenecen a dichas materias. Ambos tipos se corresponden respectivamente con los metadatos independientes y dependientes del contenido en Kashyap y Sheth. Formalmente los metadatos basados en el recurso (independientes del contenido) y los basados en la materia (dependientes del contenido) son idénticos, en el sentido de que las materias se modelan como propiedades de los recursos de información; además, todos los formatos o esquemas (*schemas*) de metainformación incluyen datos de ambos tipos.

Se podrían establecer y reflejar otras muchas clasificaciones, pero quizás la más completa, la que tiene una mayor aceptación en la literatura científica sobre este tema y que nos servirá para plantear distintas reflexiones en esta investigación, es la que se funda en la complejidad estructural de los modelos de metadatos. A pesar de que son muchos los autores que se basan en este criterio para tipificar la metainformación, sobre todo a partir del año 97, los que proponen una clasificación más sistemática en este sentido son Dempsey y Heery¹⁰³. Lo más interesante de su división de los

¹⁰³ Estos autores han expuesto esta clasificación, más o menos detallada, en casi toda su bibliografía sobre este tema. Es preciso destacar empero, que esta sistematización de tipos de metadatos la plantean en todas ellas como una mera propuesta de análisis, lejos de plantearse como una división fundamental. A pesar de no ser una clasificación pretenciosa es, desde nuestro punto de vista, la más interesante, sin ser definitiva, ya que por su propia evolución, los distintos tipos de formatos podrían cambiar de una categoría a otra, y la tendencia es que todos evolucionen hacia la banda 3 de mayor riqueza estructural y más concretamente hacia una sintaxis XML.

A tenor de las últimas tendencias de los modelos de metainformación, la clasificación de Dempsey y Heery podría parecer desactualizada en cuanto a los formatos que destacan en cada una de las bandas, ya que fue formulada en 1996-1998, sin embargo la sólida fundamentación de sus criterios de clasificación nos ha permitido adaptar, tanto la exposición como la tabla de tipificación propuesta por ellos (Tabla 8), a la realidad actual de los modelos. Dicha división de los tipos de metadatos aparece, entre otras publicaciones, en:

- Lorcan Dempsey. Roads to Desire: Some UK and Other European Metadata and Resource Discovery Projects [documento HTML]. *D-Lib Magazine*, July/August 1996, vol. 2, nº 7/8. Disponible en: <http://www.dlib.org/dlib/july96/07dempsey.html> (consultado el 18 de junio de 1998).

[cont.]

metadatos es que no se basa en un sólo criterio de caracterización, sino en varios aspectos concomitantes, que sirven para agrupar, no sólo los tipos de metadatos desde el punto de vista de los atributos que comportan, sino también los distintos modelos de metainformación. Agrupan pues, los metadatos y los esquemas de metadatos en tres tipos que ellos denominan *bandas* o *zonas*¹⁰⁴.

Para fundamentar esta clasificación y definir las distintas bandas, se basan en el criterio del crecimiento de su riqueza estructural y semántica y de su complejidad. Según esta tipología, a los metadatos registrados en la zona 3, se les presume una mayor riqueza y complicación que a aquellos situados en las bandas 1 y 2. Se apoyan pues, en un *continuum* de complejidad o en un espectro de riqueza de descripción, a lo largo del cual los metadatos son sucesivamente más completos, más estructurados, más especializados y tienen un coste de creación superior.

Así, distinguen varios tipos de metadatos agrupándolos en zonas con características comunes con relación a varios criterios, fundamentalmente: campo de aplicación, método de creación, requisitos funcionales, complejidad de designación y protocolos de búsqueda asociados a cada formato. Determinan tres zonas o bandas de metadatos¹⁰⁵, que describiremos a continuación, sin que ello sea óbice para una profundización mayor de algunos aspectos en páginas sucesivas de este trabajo.

- Lorcan Dempsey and Rachel Heery. *A Review of Metadata...* *Op. cit.*, <http://www.ukoln.ac.uk/metadata/desire/overview.rtf>, p. 8-11.

- Lorcan Dempsey and Rachel Heery. *Metadata: A Current View of...* *Op. cit.*, p. 155 y ss.

¹⁰⁴ [...] *groups the formats into bands along a continuum of growing structural and semantic richness* Lorcan Dempsey and Rachel Heery. *Metadata: A Current View...* *Op. cit.*, p. 155. Teniendo en cuenta que la categorización que realizan Dempsey y Heery está en función de una creciente complejidad, riqueza semántica, estructuración y especialización, podríamos hablar incluso de generaciones de metadatos. A pesar de que la mayoría de estos modelos de metadatos son simultáneos en el tiempo, se podría hablar de "generaciones" en el sentido de que existe una evolución estructural en cada banda señalada por estos autores.

¹⁰⁵ Rachel Heery llega a distinguir una cuarta banda, distribuyendo algunos de los formatos que aquí estudiaremos en la tercera zona, entre la banda 3 (MARC, cabeceras independientes TEI, mensajes EDI, encabezamientos SGML) y la banda 4 (donde clasifica a los formatos: ICPSR, FGDC, CIMI, [cont.]

Banda 1:

En esta zona se sitúan los datos que utilizan algunos sistemas de indización textual automatizada en Internet, que posibilitan la localización y la recuperación limitada de recursos de información. El entorno de aplicación de los metadatos comprendidos en esta banda es, o debería ser, el de los buscadores de la Web. En cuanto al modo de creación, se generan automáticamente según el propio proceso de almacenamiento de los robots indizadores, donde un software, extrae de manera automática los datos de las páginas Web basándose en las etiquetas del código HTML —en el mejor de los casos, en las etiquetas <META>, pero normalmente de la etiqueta <TITLE> y de la primera parte del documento <BODY>—.

En cuanto a la funcionalidad del registro de metadatos, es importante matizar que los modelos que podemos considerar en esta zona, sólo son útiles en el caso de la búsqueda de ítems conocidos, y no para recuperaciones genéricas por materias, porque solamente operan al nivel de copia del registro original que se ha almacenado previamente en sus bases de datos propietarias. Tampoco se puede hablar de una complejidad de designación en estos metadatos, ya que consisten únicamente en índices del texto completo cuyo contenido semántico se limita a las etiquetas HTML y a los URLs. Dempsey y Heery, señalan además, completando la caracterización, que el protocolo de búsqueda asociado a este tipo de metadatos es exclusivamente el

EAD). Los modelos de esta nueva banda, que no incluimos en esta clasificación por ser simplemente una escisión de la banda 3, se caracterizan por ser formatos de marcas altamente estructurados y por constituir estándares en materias muy específicas. Rachel Heery. *D1.1 Metadata Formats* [documento RTF]. En: *WP Study of Metadata*, issue 1.0. Bath: University, UKOLN, 23 de diciembre de 1996. Disponible en: <http://hosted.ukoln.ac.uk/biblink/wp1/d1.1.rtf>, p. 15-17 (consultado el 24 de junio de 1998).

Supone la banda más alta del *continuum* en lo que a riqueza semántica y estructuración se refiere, pero desde nuestro punto de vista, era más un rango futurible que una realidad constatada en los formatos existentes en el momento en que Heery publicaba esta división (1996). Quizás esa banda 4 sea hoy una "metabanda" que protagonicen todos los modelos basados en XML/RDF (*Cfr.* 5.2.3.).

protocolo básico de la Web, HTTP (*HyperText Transfer Protocol*) con *scripts* CGI (*Common Gateway Interface*).

Desde nuestro punto de vista, los datos clasificados en esta banda, no pueden denominarse metadatos en sentido estricto, ya que no dejan de ser datos extraídos de forma automática del propio documento y no tienen una estructuración ni una carga semántica contextual que permita, por ejemplo, la búsqueda cualificada por campos. Tampoco podemos hablar en esta zona de clasificación de un formato tipo o de un estándar de metadatos utilizado de forma pródiga por estos sistemas de indización, si bien es cierto, que quienes han implementado estos sistemas basados en la indización automatizada están tratando de desarrollar formatos simples de intercambio.

Banda 2 o banda media:

Esta zona de la clasificación, se ha denominado también *zona de recuperación de información en Internet*¹⁰⁶, y a ella pertenecen datos con una estructura sencilla y una cobertura general. Es decir, metadatos utilizados por los servicios que contienen descripciones suficientemente amplias para permitir al usuario evaluar la utilidad potencial de un recurso sin la necesidad de acceder a él, pero que no tienen un nivel de complejidad tan grande que se precisen especialistas para su creación. Esta doble caracterización (descripciones amplias y poco nivel de complejidad) se da en algunos modelos de metadatos surgidos del mundo de la informática para soportar servicios de búsqueda y directorio¹⁰⁷.

¹⁰⁶ *Discovery Zone*, en el sentido que damos en este trabajo al término *discovery*. Vid. Glosario. Lorcan Dempsey and Rachel Heery. *A Review of Metadata...* Op. cit., <http://www.ukoln.ac.uk/metadata/desire/overview.rtf>, p. 10-11.

¹⁰⁷ Algunos de estos formatos son v. gr.:

- IAFA (*Internet Anonymous FTP Archives*) es uno de los intentos más antiguos de basar la recuperación de información en Internet en modelos de metainformación. Utilizado en sistemas de indización distribuida como por ejemplo, Aliweb <<http://aliweb.emnet.co.uk>> que es uno de los
- [cont.]

El campo de aplicación de estos metadatos son los servicios de información que incorporan descripciones sencillas de los recursos: por ejemplo OCLC NetFirst (basado en su propio formato interno), o los servicios de información basados en materia (*Subject-based information gateways*¹⁰⁸) del Programa de Bibliotecas Electrónicas del Reino Unido (eLib) que utilizaban originalmente, sus propios formatos internos de metadatos o la plantilla IAFA. Estos servicios aplican, en algunos casos, criterios de selección para elegir qué es lo que describen, lo que aumenta inminentemente el coste pero mejora la calidad del servicio de información.

La función de los metadatos incluidos en esta segunda banda, es la posibilidad de sustituir a la fuente principal a la que describen, de tal forma que los datos que incluyen son lo suficientemente descriptivos para que el usuario evalúe la utilidad

buscadores más antiguos de la WWW. Desarrollado por Nexor y mantenido por Emnet, este servicio de búsqueda se ha convertido también en un portal <<http://www.aliweb.com>>, donde además de otros servicios, permite realizar búsquedas acotándolas a los campos definidos en el modelo de metadatos IAFA. Para más información sobre el funcionamiento, etc. de esta plantilla de metainformación, *Vid.* David Beckett. IAFA Templates in Use as Internet Metadata [documento HTML]. En: *International World Wide Web Conference (4. 1995. Boston)*, rev. 15 de diciembre de 1995. Disponible en: <http://www.w3.org/Conferences/WWW4/Papers/52> (consultado el 10 de febrero de 2001).

- WHOIS ++, protocolo de los servicios de directorio en Internet. WHOIS++ junto con IAFA constituyen la base del proyecto ROADS <<http://www.ilrt.bris.ac.uk/roads>> utilizado por muchas de las *subject gateways* del proyecto eLib que ha desarrollado un software propietario de indización de la WWW basado en metadatos.
- RFC 1807 (*A Format for Bibliographic Records* <<http://www.ietf.org/rfc/rfc1807.txt?number=1807>>). En realidad este formato es un conjunto de recomendaciones más que un verdadero estándar de metadatos, que tratan de definir un modelo para enviar por correo electrónico informes técnicos o registros bibliográficos.
- Una excepción dentro de este tipo de formatos es el caso del Dublin Core, un acervo de elementos de metadatos sencillo que ha surgido de una iniciativa en el contexto bibliotecario, y que hoy por hoy es uno de los formatos que han adquirido mayor desarrollo, evolución y aceptación. *Vid.* 5.2.1.2.

¹⁰⁸*Vid.* 8.2.2.

potencial que puede tener el recurso para sus necesidades de información, sin que para ello tengan que consultarlo, soportando incluso, servicios de directorio¹⁰⁹.

En cuanto al modo de creación, no se precisa una formación especializada ya que pueden generarse bien de forma manual, o a través de una validación manual de las descripciones automáticas, incluyendo una variedad de atributos descriptivos y de otros tipos. El encargado de crear este tipo de metadatos, puede ser el propio administrador de la Web o una agencia central en su propia base de datos, sin necesidad tampoco de tener conocimientos específicos en el dominio científico en que se utiliza (*Vid.* 4.2 y 4.3).

Por otra parte, no son formatos que contengan una estructura interna muy elaborada, pero a pesar de su aparente sencillez, tienen dificultades para representar estructuras jerárquicas u otros objetos yuxtapuestos, es decir, son formatos abocados a la descripción de objetos discretos. Esto implica que los modelos de metadatos de esta zona no son adecuados para la representación de DLOs complejos o mixtos, por ejemplo texto, imagen y sonido, etc., ni documentos con una estructura jerárquica complicada, como por ejemplo un tesoro hipertextual.

En lo relativo a la complejidad de designación, a pesar de no tener estructuras internas muy elaboradas, contienen el nivel de designación suficiente para la

¹⁰⁹ Esto tiene relación con el concepto de sustituto documental (*Vid.* Glosario, *surrogate*, *subrogate*) que toma un nuevo cariz en el contexto de los metadatos. Partiendo el concepto de sustituto digital, la diferencia entre los servicios que implementan metadatos de la banda 1 y la banda 2 desde el punto de vista de la recuperación de información, sería similar en cierta forma, a la que puede existir, en el contexto de la documentación tradicional, entre un registro de una bibliografía o de un catálogo manual y el registro de una base de datos comercial. Esta comparación puede parecer un poco excesiva, y requeriría una mayor matización, sin embargo sólo queremos introducir que, en el ámbito de comparación de modelos de metadatos, los criterios de evaluación de su utilidad para la recuperación de información son similares a los que adoptamos en el contexto tradicional, y el nivel en que una descripción es capaz de sustituir, o al menos de anticipar, la utilidad de una publicación, puede ser uno de ellos.

búsqueda por algunos campos. Sin embargo, como ya hemos apuntado, la descripción se realiza sobre objetos discretos y no recoge ni las relaciones existentes entre ellos, ni las implicaciones jerárquicas. A pesar de todo, gracias a una construcción y contenido similares, y al relativo bajo coste de creación de sus registros, estos formatos de metadatos son los candidatos más adecuados para describir los recursos de servicios de información donde se entremezclan distintos tipos de medios y dominios.

En relación con los protocolos de búsqueda asociados, se señala el auge creciente que están teniendo en Internet los servicios de indización y búsqueda distribuida y los directorios, especialmente: WHOIS++, LDAP¹¹⁰ (*Lightweight Directory Access Protocol*) y Dienst; en algunos casos también se usa el protocolo Z39.50, sobre todo a partir de la evolución de éste en los dos últimos años (2000-2001).

En el proceso de maduración de propuestas de metadatos al que asistimos, en nuestra opinión desde 1999, están provocando que los estándares de datos se adapten a las tendencias emergentes de búsqueda y recuperación de información en la Red, o mejor dicho, que las tendencias de búsqueda se adapten a estos nuevos esquemas de metainformación. Sin embargo, podemos prever que surgirán otros esquemas alternativos normalizados formalmente, o que algunos de los modelos comprendidos en esta zona se optimicen como demuestran por ejemplo, los avances del Dublin Core (desde Helsinki 1997¹¹¹) y la co-evolución con RDF de este formato. De

¹¹⁰ Para corroborar el desarrollo del uso de este protocolo y la utilización de servidores LDAP y bases de datos basadas en XML, Cfr. Ben. Y. Zhao, Anthony D. Joseph. *Xset: A Lightweight Database for Internet Applications* [documento PDF]. Berkeley: University of California, Computer Science Division, rev. 5 de junio de 2000. Disponible en: <http://www.cs.berkeley.edu/~ravenben/xset/html/xset-saint.pdf> (consultado el 25 de octubre de 2000).

¹¹¹ Sobre los avances concretos del V taller del Dublin Core y la descripción formal para codificar los metadatos basados en este modelo cualificado de la banda 2 en una nueva expectativa formal (*Resource Description Framework*). Vid. Stuart Weibel and Juha Hakala. DC-5. The Helsinki [cont.]

momento, este rango de metadatos tiene a su favor la semejanza de construcción y por ello, la aptitud para el intercambio, conversión e interoperabilidad entre modelos.

Banda 3 o documentation band

En esta tercera zona o "generación" de metadatos encontramos formatos más completos y detallados, mejor estructurados y normalmente desarrollados en el seno disciplinas científicas concretas que quieren contar con los requisitos funcionales específicos para un uso eficaz de sus materiales digitales. La tercera banda, abarca algunos sistemas muy completos para la descripción, no sólo de objetos, sino también de colecciones de objetos, prestando atención tanto a la estructura de la información, como a su contenido.

Dentro de las características que comparten los formatos situados en esta zona, podemos subrayar su riqueza funcional (mayor que en las dos categorías anteriores), así como su capacidad para expresar gran variedad de relaciones semánticas y formales. Destaca también el hecho de que en esta banda, los metadatos suelen ser parte de un marco semántico más amplio que incluye un lenguaje de marcas para el diseño del contenido de los documentos. Además, los metadatos de última generación circunscritos a esta banda están basados originalmente en el metalenguaje SGML y, ahora cada vez más, estructurados en XML; en otras ocasiones responden a formatos individuales como es el caso del ICPSR (*Inter-University Consortium for Political and Social Research*) o el modelo desarrollado para el control de la información geoespacial, CSDGM del Comité Federal de Datos Geográficos (FGDC). Inclusive se contempla dentro de esta zona el formato MARC, es decir, el formato más generalizado en el contexto bibliotecario como base de la catalogación automatizada. Otra de las particularidades importantes de los modelos situados en

Metadata Workshop: A Report on the Workshop and Subsequent Developments [documento HTML]. *D-Lib Magazine*, February 1998, vol 4, nº 2. Disponible en: <http://www.dlib.org/dlib/february98/02weibel.html> (consultado el 17 de julio de 1998).

esta última zona es la necesidad de un conocimiento especializado del formato en cuestión para la creación (que puede ser manual) y mantenimiento de metadatos.

En cuanto a la funcionalidad de los registros construidos según estos modelos, es preciso señalar que los formatos que incluyen Dempsey y Heery en esta banda posibilitan descripciones prolijas que pueden usarse tanto para la localización de recursos conocidos, como para la búsqueda y recuperación genérica, son aptos tanto para la navegación como para establecer gran variedad de relaciones entre registros a distinto nivel.

La complejidad de designación no es muy grande y los usuarios especializados pueden buscar información usando directamente este formato. Para que estos metadatos sean más transparentes en las herramientas generales de búsqueda pueden exportarse como datos de un formato más simple, por ejemplo, a aquellos situados en la banda 2.

Sólo resta mencionar, que el protocolo de búsqueda más importante y generalizado, asociado a esta banda de formatos, es Z39.50. Desde finales de 2000 y teniendo en cuenta la adecuación de este protocolo para el acceso a las colecciones digitales, se reconoce cierta integración de Z39.50 con estándares de metadatos basados en XML; por ejemplo, a través de la implantación de pasarelas Z-Web, el uso de XML como sintaxis de registro, o la incorporación del DC para la búsqueda. En los últimos dos años, se ha barajado también la posibilidad de redefinir la norma Z39.50 de tal suerte que se asegure su integración con la nueva forma de estructuración de la información Web, o bien la creación de un nuevo protocolo de intercambio bibliográfico basado en XML que soporte estándares de metadatos. Finalmente, en la reunión del ZIG de octubre de 2001 celebrada en la ciudad británica de Boston¹¹², se decidió definir la ZNG (*Z39.50 Next Generation*) que

¹¹² Para una visión más profunda de los distintos temas debatidos en esta reunión, Vid. <http://lcweb.loc.gov/z3950/agency/zig/meetings/uk2001/agenda.html>

soporte nuevos formatos de metainformación como DC/RDF, ONIX (*Online Information eXchange*), OAI/MARC- XML, etc., así como un nuevo modelo de Z39.50 orientado a objetos (ZOOM), lo que supone una revisión inaplazable de la norma Z del año 1995¹¹³.

Recogemos a continuación un cuadro resumen de la clasificación que plantean Dempsey y Heery¹¹⁴ en una publicación realizada como parte de su investigación en el proyecto DESIRE¹¹⁵. Sintetizan la tipología que defienden de la siguiente forma:

¹¹³ ANSI/NISO Z39.50-2001. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification [documento PDF]. Bethesda, Maryland: NISO, rev. 21 de noviembre de 2001. Disponible en: <http://www.loc.gov/z3950/agency/revision/part1.pdf> y <http://www.loc.gov/z3950/agency/revision/part2.pdf> (consultado el 29 de noviembre de 2001).

¹¹⁴ Lorcan Dempsey and Rachel Heery. *A Review of Metadata...* Op. cit., <http://www.ukoln.ac.uk/metadata/desire/overview.rtf>, p. 8. La columna en **sombra** la hemos añadido nosotros a tenor de nuevos estándares de metadatos asimilables a alguna de las bandas, o bien teniendo en cuenta matizaciones, como en el caso de la banda 1, que parece más adecuado definir de forma genérica etiquetas <META> del HTML, que los pseudo-metadatos que puedan asimilarse a sistemas de búsqueda concretos.

¹¹⁵ DESIRE (*Development of a European Service for Information on Research and Education*) es un proyecto europeo consolidado por el programa *Telematics for Research Sector* dentro del IV Programa Marco para la investigación y el desarrollo tecnológico de la UE. Este proyecto se centra en la tecnología Web y en la creación de servicios de información piloto que puedan beneficiar a los investigadores europeos. En este proyecto participan distintas instituciones europeas; concretamente Lorcan Dempsey y Rachel Heery, pertenecen al grupo de trabajo 3 (WP3) sobre Recuperación de recursos e indización, en representación del UKOLN (*UK Office for Library and Information Networking*). Para más información sobre el proyecto DESIRE, Vid. <http://www.nic.sufnet.nl/surfnet/projects/desire/desire.html> (consultado el 6 de julio de 1998); <http://www.desire.org> (consultado el 5 de febrero de 2001).

	Banda 1		Banda 2		Banda 3	
Características	Formatos simples Sistemas propietarios Indización full-text		Formatos estructurados Estándares <i>de facto</i> Estructura de campos		Formatos ricos (<i>XML</i>) Estándares internac. Etiquetas elaboradas	
Formatos	Lycos AltaVista Yahoo, Etc.	Etiquetas <META> en HTML	DC/DCMI IAFA RFC 1807 SOIF LDIF	EdNA AGLS etc.	ICPSR CIMI EAD TEI MARC	DC-AP DIG35 etc.

Tabla 8. Tipos de metadatos en bandas (Dempsey y Heery).

Sutton hace su propia interpretación de este *continuum*, dando una mayor relevancia a la funcionalidad de los metadatos, tipificándolos según el estereotipo de cada banda. Su interpretación arranca de la simplicidad de las descripciones que existen en los motores de búsqueda, a la complejidad y riqueza descriptiva de los registros bibliográficos en MARC. Según Sutton¹¹⁶, las funciones de los metadatos a lo largo de este *continuum* van desde la simple localización de ítems de información, a una descripción completa y compleja que incluye información no contenida explícitamente en el DLO descrito, pasando por una situación intermedia de recuperación de información (Tabla 9). Esa situación intermedia, protagonizada principalmente por el formato Dublin Core con una sintaxis RDF/XML, es la que trataremos de resaltar como nivel de metadatos idóneo para la gestión y recuperación de información en bibliotecas digitales.

¹¹⁶ Stuart A. Sutton. Conceptual Design and Deployment of Metadata Framework for Educational Resources on the Internet. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, n° 13, p. 1185.

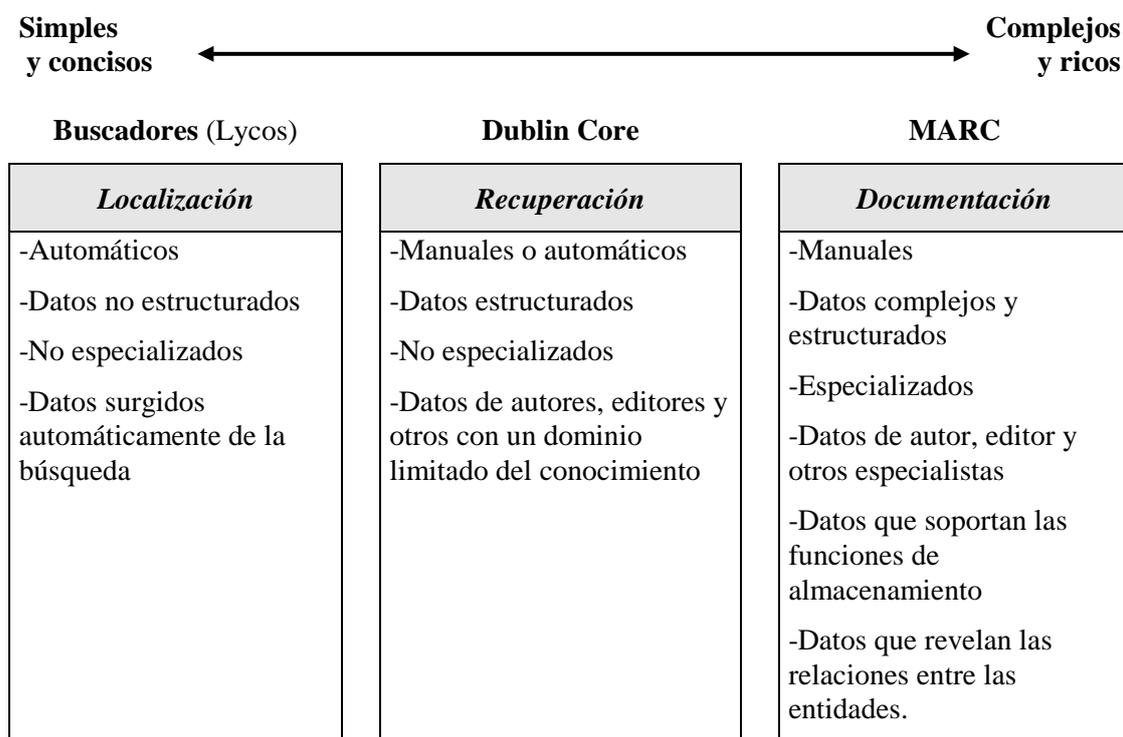


Tabla 9. Interpretación del *continuum* de metadatos (Sutton)

La división de los metadatos en bandas ha dado lugar a otras interpretaciones¹¹⁷ basadas en las reflexiones de Dempsey y Heery, y en algunos casos también, protagonizadas por ellos¹¹⁸, que plantean una clasificación en tres tipos de metainformación que se corresponden en lo sustancial con las tres zonas que venimos mencionando. Este es el caso de la clasificación en: metadatos desestructurados (banda 1), estructurados (banda 3) y semi-estructurados (banda 2). Los metadatos desestructurados corresponden a la indización Web carente de semántica y que facilita la búsqueda en texto libre; los formatos de una estructura intermedia tienen

¹¹⁷ P. ej., Andrew Large, Lucy A. Tedd and R. J. Hartley. *Information Seeking in the Online Age: Principles and Practice*. London, etc.: Bowker Saur, 1998, p. 137.

¹¹⁸ Lorcan Dempsey, Rosemary Russell, Rachel Heery. In at the Shallow End: Metadata and Cross-domain Resource Discovery. En: *Discovering Online Resources Across the Humanities: A Practical Implementation of the Dublin Core*. Paul Miller and Daniel Greenstein, eds. Bath: UKOLN, 1997, p. 64-65.

una sintaxis de pares de atributo-valor con algunas particularidades semánticas como el modelo DC y sirven para realizar una búsqueda por campos; y finalmente, los metadatos altamente estructurados, suelen utilizarse en dominios específicos de aplicación y tienen una fuerte carga sintáctica y semántica basada en lenguajes de marcado como SGML o XML, y su utilización puede ir desde la recuperación hasta la preservación de información electrónica.

Este mismo aspecto del *continuum* de complejidad, o más bien del *continuum* estructural, es un tema recurrente en el ámbito de los metadatos y fue uno de los aspectos centrales del cuarto *Workshop del Dublin Core* celebrado en Canberra en marzo del 97 donde se discutió acerca de la riqueza estructural de los calificadores enfrentando dos posturas, la de los minimalistas y la de los estructuralistas defendiendo unos, la simplicidad del modelo al representar los campos mínimos, y otros, el nivel de especificidad semántica que aportaban representaciones calificadas para las necesidades de un área de información particular¹¹⁹.

Para finalizar este apartado e iniciar al mismo tiempo algunas reflexiones que retomaremos en el capítulo siguiente, planteamos a continuación una clasificación basada en la idiosincrasia del propio proceso de creación de los metadatos, considerado distintos criterios o aspectos que confluyen en la asignación de metadatos (*Vid.* Capítulo 4). Así, podemos identificar una sencilla, aunque valiosa tipología que nos servirá para completar la caracterización de estos elementos, al mismo tiempo que para establecer una nueva clasificación.

¹¹⁹ Stuart Weibel, Renato Ianella, Warwick Cathro. The 4th Dublin Core Metadata Workshop Report [documento HTML]. *D-Lib Magazine*, June 1997, vol. 3, n° 6. Disponible en: <http://www.dlib.org/dlib/june97/metadata/06weibel.html> (consultado el 16 de junio de 1998).

En el capítulo 5, al hablar del Dublin Core (5.2.1.2.), profundizaremos en los distintos talleres o *workshops* celebrados en torno a dicho modelo de metadatos. El hecho de traerlo aquí ahora, se debe a la importancia y desarrollo de dicho esquema de metainformación, así como para ilustrar que el factor de la complejidad estructural no sólo sirve para clasificar los distintos esquemas de metadatos, sino también para generar escuelas o tendencias en el seno de un mismo modelo.

- Atendiendo a la fuente y al momento de creación, podemos distinguir por un lado, metadatos internos, generados para un DLO en el momento de su creación y/o digitalización, y por otro, metadatos externos, relacionados con un documento como objeto, pero creados *a posteriori* por alguien distinto al agente original que crea la información digital¹²⁰.
- Según el método concreto de creación, podemos distinguir metadatos generados automáticamente por el ordenador (como por ejemplo los índices de palabras clave), y metadatos creados manualmente en la cabecera del documento. En esta misma línea, podríamos señalar otro tipo de metadatos creados manualmente pero utilizando una herramienta informática, una plantilla o cualquier aplicación que facilite su creación.
- Si tenemos en cuenta el nivel de especialización de los responsables en el proceso de asignación de metadatos, sean estos manuales o automáticos, podemos identificar también dos tipos de metadatos: aquellos creados por personas que no son especialistas temáticos, ni especialistas en documentación (el caso de los metadatos creados por el propio autor de una página personal); y aquellos creados por expertos, bien en la materia de la que trata el DLO en cuestión, bien en información y documentación. A este último tipo correspondería, la creación de un registro en MARC o de la asignación especializada de metadatos, por ejemplo, en una biblioteca digital de información de patrimonio, que utiliza tanto las estructuras específicas (v. gr. el modelo del Consorcio para el Intercambio Automatizado de Información de museos (CIMI), como un lenguaje controlado adecuado a la materia (p. ej. el tesoro de Arte y Arquitectura (AAT) del Getty).

¹²⁰ En el ámbito de las bibliotecas digitales, los metadatos internos responden a las bibliotecas productoras y poseedoras de su colección digital, mientras que los metadatos externos son la forma habitual de metadatos en las *subject gateways* formadas por colecciones virtuales. *Cfr.* 8.2.1. y 8.2.2.

- En relación con este último aspecto, podemos utilizar otro factor de clasificación, el control semántico de los datos. Así, podemos distinguir: metadatos controlados que responden a un vocabulario estándar, lista de autoridad o tesauro, y metadatos no controlados que no responden a ninguna de estas herramientas documentales de organización del conocimiento.
- Según la forma de almacenamiento se pueden distinguir, utilizando la propia terminología de la DCMI¹²¹: metadatos embebidos en el propio documento, cuando estos se almacenan con el propio objeto que describen; y metadatos separados del documento (*stand-alone metadata*), aunque como veremos, la casuística de almacenamiento, en virtud a su recuperación ulterior, es bastante más compleja y diversa.
- Por el nivel de estructuración, podemos advertir una gradación de la metainformación que va desde metadatos altamente estructurados (como el MARC, las cabeceras TEI, etc.), a metadatos no estructurados (como anotaciones que no responden a una estructura previsible, pero que dan información sobre el documento).
- Según el nivel de descripción, y en estrecha relación con lo que sería la unidad documental base (en este caso un DLO) tomada para la descripción, encontramos: metadatos de colección, que realizan un registro a nivel colección, y metadatos individuales, que toman como referencia una unidad mínima digital, por ejemplo una imagen, y que generalmente están incluidos dentro de una colección o de un DLO más complejo.
- Finalmente, si atendemos a la perdurabilidad, podemos destacar cuatro tipos: metadatos estáticos, que no suelen cambiar una vez creados (título, fecha de

¹²¹ Mary S. Woodley. Glossary [documento HTML]. DCMI, 12 de abril de 2001, rev. 24 de abril de 2001. Disponible en: <http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml> (consultado el 8 de junio de 2001).

creación, etc.); metadatos dinámicos, que pueden cambiar con el uso y la manipulación del DLO (por ejemplo, datos relativos a la resolución de las imágenes); metadatos de larga duración que aseguran que el objeto de información digital siga siendo accesible y, metadatos de corta duración, normalmente los de tipo operacional (como los datos relativos a la administración del documento).

En cualquier caso, y a pesar de lo atinado de algunas de las clasificaciones que hemos destacado, son muchas más las que podríamos hacer de los metadatos. Pero la finalidad de este trabajo va más allá de la mera categorización de estos *datos sobre los datos*.

Con todo lo que hemos dicho a lo largo de este capítulo podemos afirmar que los metadatos son un constructo complejo que sirve para identificar, localizar, gestionar y acceder a los recursos electrónicos. A esto hay que añadir que la aplicación de metainformación es costosa tanto desde el punto de vista de la creación como del mantenimiento, por ello, a lo largo de los capítulos siguientes trataremos de razonar su uso y de defender nuestra propia hipótesis de aplicabilidad de los mismos.

CAPÍTULO 4:

EL PROBLEMA DE LA ASIGNACIÓN DE METADATOS

Hasta ahora, nos hemos centrado en el qué y el para qué de los metadatos destacando su valor como tecnología complementaria y potencial para el problema de la recuperación relevante en Internet. Pero ¿quién hará el trabajo de crear los metadatos necesarios para la Web? Trataremos pues a continuación, una de las cuestiones fundamentales en torno a este tema, que tiene además relación directa con el trabajo bibliotecario: determinar el proceso de creación, esto es, establecer el *cuándo, quién y cómo* de la asignación de metadatos.

Con esta finalidad, a lo largo de este capítulo compararemos, en primer lugar, el proceso de creación de metadatos con la actividad catalográfica tradicional partiendo, para ello, de la idiosincrasia de la información Web y de la perspectiva del concepto de metainformación en sentido estricto, teniendo presente el reto de la recuperación de información electrónica. Después haremos hincapié en la discusión acerca de quién es el agente que debe establecer estos datos sobre los datos: ¿el propio autor del documento electrónico? ¿un "catalogador" especializado que revise y describa cada documento Web posteriormente a su creación? o por el contrario, ¿podemos confiar la asignación de metadatos a un agente automatizado? Finalmente, trataremos también aquí, aspectos sobre cómo dotar a los documentos Web de un registro de metadatos, para ello se analizan algunas de las herramientas y aplicaciones que se prodigan en este sentido y que hemos seleccionado de un conjunto amplio de ellas.

La proliferación de páginas Web se debe a múltiples razones, entre las que podemos destacar: el alcance global de este medio de edición electrónica, la facilidad

de creación y mantenimiento de documentos, la escasa complejidad del lenguaje HTML así como la existencia de herramientas que permiten editar y modificar de forma sencilla las páginas Web, el bajo coste de este tipo de publicaciones, la existencia de servidores gratuitos donde alojar las páginas, el hecho de que no estén establecidos unos criterios mínimos de calidad de los documentos para su publicación, etc.

No obstante, no es habitual que las páginas incluyan la etiqueta <META> donde se introducen los metadatos en la sintaxis convencional de la WWW, es decir, en HTML¹. En el mejor de los casos, los autores incluyen la etiqueta <META NAME="keywords"> y <META NAME="description"> (que supuestamente

¹ Como ejemplo de la poca utilización de metainformación en páginas Web, es interesante mencionar los muestreos realizados por: José Antonio Merlo Vega y Ángela Sorli Rojo. El uso de metainformación en los webs de las bibliotecas españolas. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao)*. Op. cit., p. 154-164. En este estudio se analiza la utilización de etiquetas <META> relativas a título, autor, descripción y palabras clave en una muestra de 165 páginas Web pertenecientes a otras tantas bibliotecas españolas (60 universitarias, 50 públicas y 55 especializadas), utilizando para el análisis la plantilla DC-Dot —Vid. 4.3.2.1—. De las páginas Web analizadas sólo el 7,27% de las bibliotecas utilizaba las cuatro metaetiquetas.

En un estudio semejante presentado también en las mismas jornadas, Vidal y Salvador estudian el uso de etiquetas <META> en las páginas Web de bibliotecas y centros de documentación pertenecientes a RedIris. En este caso varía la metodología de análisis, clasificando los metadatos utilizados en 4 tipos: A) descripción documental básica que incluye una etiqueta; B) etiquetas *description* y *keywords*; C) registros más completos que añaden además la fecha, lengua, organización, etc. y D) etiquetas del sistema Dublin Core), y la población analizada (156 sedes o páginas Web pertenecientes a unidades de información de RedIris), pero los resultados son parecidos: un 70,5% de los centros analizados no utiliza metadatos para la descripción de páginas Web, dentro de los cuales sólo un 2,6% utiliza el sistema DC. Francisco Javier Vidal Bordés, José Antonio Salvador Oliván. La implementación de metadatos y Dublin Core en sedes y páginas Web de bibliotecas y centros de documentación de universidades y centros de investigación de la Red Iris. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao)*. Op. cit., p. 197-209.

En el transcurso de esta investigación hemos podido analizar, de forma cualitativa en nuestro caso, infinidad de recursos electrónicos, y los resultados son parecidos. La utilización del software Metabrowser —Vid. *infr.* 4.3.2.2., Fig. 41— en su función de navegador nos ha permitido ver que las páginas a las que accedíamos como recursos de información para la realización de esta tesis (fundamentalmente sobre metadatos y/o bibliotecas digitales) tampoco utilizan sistemáticamente los modelos que describen o encomian en el cuerpo de la página.

tienen en cuenta algunos de los principales buscadores de Internet para la indización), bien a través de los menús de los editores WYSIWYG², o de aplicaciones y plantillas específicas creadas a tal efecto³. Pero lo más habitual es que estas herramientas no se utilicen y de hacerse, el usuario no tiene conciencia estructural de estar asignando metainformación a sus páginas.

Teniendo en cuenta todo esto, otro aspecto fundamental que deberíamos analizar es la cuestión de a qué asignar metadatos, qué tipo de documentos digitales entendidos como objetos (DLOs) deberían llevar ese registro sobre la información que contienen. Sin embargo, como anticipamos en el capítulo 2, hemos obviado premeditadamente este tema porque su tratamiento aquí obligaría a profundizar expresamente en otro aspecto fundamental: la evaluación de recursos Web⁴, que

² WYSIWYG: *What You See Is What You Get*. Son sistemas finales de edición cuyo acrónimo quiere decir: lo que ves es lo que obtienes. Por ejemplo Netscape Composer, una de las herramientas de edición de páginas Web más sencilla y utilizada, en tanto que está incluida en la *suite* de programas del navegador Netscape, a partir de su versión 3.0 Gold, tiene una opción para añadir las etiquetas <META> en las propiedades de la página, sin embargo, como indicamos en el texto arriba, no es frecuente que se complete esa información, incluso aunque se utilicen herramientas WYSIWYG más sofisticadas de edición de páginas Web.

³ En el apartado 4.3. de este capítulo se analizarán estas plantillas, además de otras herramientas más complejas estrictamente vinculadas a los metadatos, así como otras aplicaciones y plantillas (*templates*) sencillas concebidas explícitamente para lo que se denomina, cada vez más, "promoción de la Web". *Vid. infr.*

⁴ Sobre este aspecto Lluís Codina plantea una metodología muy interesante para la evaluación de recursos digitales donde propone 14 criterios de evaluación divididos en: *micronavegación* (calidad y volumen de la información, autoría: responsabilidad y solvencia, legibilidad y ergonomía, navegación y representación de la información, recuperabilidad, interactividad/computabilidad, velocidad de descarga y servicios adicionales) y *macronavegación* (luminosidad, calidad de los enlaces, actualización de los enlaces, descripción, selección y evaluación, visibilidad, y autodescripción). Este autor habla además de la descripción de recursos digitales y de metadatos como un aspecto dentro del procedimiento que deben seguir las agencias de información (*information gateways*, *Vid. Glosario*). Lluís Codina. Evaluación de recursos digitales en línea: conceptos, indicadores y métodos. *Revista Española de Documentación Científica*, 2000, vol. 23, nº 1, p. 9-44. Codina además alude a un estudio de la Universidad de Georgia donde se encontraron 509 posibles criterios de evaluación <<http://itech1.coe.uga.edu/faculty/gwilkinson/webeval.html>> [sic.]. Gene L. Wilkinson, Kevin M. Oliver, Lisa T. Bennett. *Evaluating the Quality of Internet Information Sources* [documento HTML]. Georgia: Department of Instructional Technology, University of Georgia, 20 de mayo de 1997. [cont.]

excedería los objetivos de esta investigación. A pesar de ello, sí reflejaremos algunas reflexiones en este sentido.

4.1. De la catalogación a la asignación de metadatos: reflexiones en torno al proceso técnico de los recursos electrónicos

Si bien, como hemos visto en el capítulo 3, existe una tendencia teórica que asimila la creación o asignación de metadatos a la catalogación (y así encontramos expresiones como *catalogación de recursos electrónicos* e incluso *catalogación de Internet*⁵), en este apartado trataremos de dirimir entre las concomitancias de ambos procesos, ajustándonos a la necesidad que impera en un nuevo entorno informativo digital y en la nueva gestión global de la información; discutiendo empero la alternativa de la catalogación tradicional de recursos electrónicos, que defienden

Disponible en: <http://itech18.coe.uga.edu/faculty/gwilkinson/webeval.html> (consultado el 3 de abril de 2001).

Junto a todas las razones para evaluar recursos digitales que se esgrimen en los documentos citados, debemos añadir que la evaluación es un paso indispensable antes de realizar la descripción de un recurso, equiparable al proceso de selección en las bibliotecas tradicionales. Se deben asignar metadatos a recursos de calidad susceptibles de formar parte de una biblioteca digital, sin que ello obste para que los autores/editores de recursos Web incorporen metaetiquetas en aras a la promoción de sus Webs y a la recuperación a través de herramientas de búsqueda genéricas.

⁵ Bajo estas denominaciones destacan las múltiples tentativas y proyectos de OCLC, como por ejemplo: los coloquios que han organizado sobre este tema: *OCLC Internet Cataloging Colloquium* (1996. San Antonio). OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/xu.htm> (consultado el 19 de abril de 2000); las publicaciones monográficas dirigidas a la utilización del formato MARC y del campo 856 para la localización electrónica y acceso de documentos electrónicos catalogados, como: *Cataloging Internet Resources: A Manual and Practical Guide* [documento HTML], Nancy B. Olson, ed. 2nd ed. OCLC, rev. 28 de julio de 1998. Disponible en: <http://www.oclc.org/oclc/man/9256cat/toc.htm> (consultado el 1 de abril de 2001). En esta misma línea se inserta también, entre otras, la publicación *Journal of Internet Cataloging*. Esta asimilación puede deberse, como decíamos en el capítulo anterior, al término torpe y a la definición sibilina de los metadatos. En la línea argumental sobre la poca eficacia designativa del término metadatos, nos parece especialmente elocuente la que da Dempsey de "torpe" a este término y que da lugar a la asimilación de los metadatos a la catalogación. *Cfr.* Capítulo 3, notas 13, 15.

muchos autores y que corroboran algunos proyectos de aplicación del formato MARC y de las reglas de catalogación.

El proceso de asignación de metadatos podríamos definirlo como el "arte" de añadir a la estructura de un documento digital, la cabecera que describe el contenido, la propiedad intelectual, la tipología, la materia, etc. del documento, en definitiva, la actividad que consiste en extraer y añadir información sobre la información publicada en la Web en aras a su ulterior recuperación. Se trata pues, de un "arte" de cariz técnico, ya que para añadir esas metaetiquetas se requiere cierta destreza y conocimiento de los metalenguajes SGML/XML o de lenguajes de marcado específicos como HTML, en que esté realizada la publicación electrónica, así como del estándar o modelo de metadatos aplicable a tal efecto y del sistema de búsqueda utilizado para la recuperación según un determinado esquema de metainformación.

Por otra parte, la catalogación tradicional, también podría entenderse como un "arte" técnico, ya que requiere el manejo de unas normas bibliográficas así como de la aprehensión de todo un sistema catalográfico local. La catalogación ha servido a las bibliotecas en un doble sentido: por un lado como una forma de proporcionar acceso a una colección de recursos de conocimiento, y por otro como una manera de gestionar una colección de productos impresos. En un entorno informativo disperso, descentralizado y global como el que impone la WWW, las funciones de la catalogación se desvanecen. Quizás los metadatos asuman estas funciones (proporcionar acceso y gestionar una colección —digital, e incluso virtual, como veremos—) en el ámbito de la información electrónica pero, desde nuestro punto de vista, la función de los metadatos trasciende a la catalogación o mejor dicho, es distinta. Con esto no estamos contradiciendo la tendencia conceptual que asimila la catalogación a los metadatos, sino más bien que la mayor parte de los metadatos no se pueden considerar catalogación; y no lo son, porque las necesidades descriptivas de los recursos electrónicos son diferentes.

Según la descripción que hace la Biblioteca del Congreso de los Estados Unidos de su sistema de catalogación:

Cada registro bibliográfico contiene una descripción de un ítem como forma de identificarlo y distinguirlo de otros. El registro contiene los medios para proporcionar acceso a través de diferentes puntos, incluido el autor y otras entidades asociadas al trabajo o a través de relaciones con otros trabajos del catálogo, título (incluido el título uniforme para ordenar los trabajos publicados con diferentes títulos), series y materias. El acceso por materias se realiza a través de encabezamientos de materia y clasificaciones [...]»⁶.

Si partimos de esta concepción de la LC, podemos entender también que los metadatos son la descripción/representación de un documento, independientemente del soporte o de que sean DLOs de Internet, y por tanto, necesitarán completar las tres funciones implícitas en la catalogación convencional: 1) identificar excepcionalmente cada objeto de información en una colección determinada; 2) proporcionar distintas formas de acceso y recuperación de cada ítem; y 3) relacionar la información contenida en cada documento con otros documentos, informaciones y/o conocimiento.

Banerjee⁷, de forma análoga a las funciones tradicionales de la catalogación, explica que la catalogación de recursos electrónicos consiste en tres pasos básicos: 1) *proporcionar la información descriptiva que identifica el trabajo*; 2) *determinar los puntos de acceso necesarios para la recuperación del registro en el catálogo*; y 3) *almacenar la forma en la que se puede acceder al trabajo*. El Zaïm y Tellier⁸, de

⁶ *Modes of Cataloging Employed in the Cataloging Directorate* [documento HTML]. Washington: Library of Congress, 27 de enero de 1997. Disponible en: <http://lcweb.loc.gov/catdir/catmodes.html> (consultado el 31 de octubre de 2000).

⁷ Kyle Banerjee. Describing Remote Electronic Documents in the Online Catalog: Current Issue. *Cataloging and Classification Quarterly*, 1997, vol. 25, nº 1, p. 16-17.

⁸ Adel El Zaïm et Sylvie Tellier. L'avenir des formats de communication: HTML, SGML et formats bibliographiques: de l'information à la métainformation [documento HTML]. En: *International Conference on The Future of Communication Formats (1996. Ottawa)*. Hull: Banque Internationale d'information sur les États francophones, rev. 23 de febrero de 1996. Disponible en: <http://www.acctbief.org/avenir/crim.htm> (consultado el 10 de abril de 2001).

manera más rigurosa, dicen que la representación documental que supone la metainformación se realiza a tres niveles: *signalético, analítico y referencial*.

- La descripción signalética consiste en extraer los elementos generales de un documento (autor, título, editor, etc.) para señalar su existencia. Esta descripción signalética, podría asimilarse a lo que tradicionalmente denominamos análisis formal del documento.
- La descripción analítica consiste en tomar conciencia del contenido de un documento con la ayuda del título, del sumario, de las cabeceras de los capítulos, o haciendo una lectura diagonal, de tal forma que se condense el contenido y las materias del documento en un resumen o en una lista de palabras clave. Este nivel de descripción es lo que tradicionalmente llamamos análisis de contenido.
- Y finalmente, señalan una función referencial, que se produce al relacionar descripciones signaléticas en otro documento, por ejemplo cuando se realizan bibliografías o listas de referencias y, en el caso de la información electrónica, cuando se establece una red de citas a través del hipertexto.

Estas triples descripciones procedimentales que consignamos (la de la catalogación convencional —LC—, la de la catalogación de recursos electrónicos —Banerjee— y la de la asignación de metadatos — El Zaïm y Tellier—) se pueden aplicar tanto al proceso de descripción bibliográfica y análisis de contenido tradicionales, como a los nuevos proceso de adopción y adaptación de esquemas de metadatos. Sin embargo, debemos destacar que las particularidades de la información Web, a la que representan o tratan de sustituir⁹ los distintos metadatos, difieren de las características de la información tradicional. Por ello, y antes de adentrarnos en el problema del procesamiento de los recursos electrónicos para su recuperación, parece

⁹ Vid. Capítulo 3, nota 109. Vid. Glosario, *surrogate*, *subrogate*.

oportuno —emulando la metodología de Heery¹⁰ y la nuestra propia¹¹— plantear una reflexión sobre esas características específicas de la información electrónica que justifican, en cierta medida, que un registro de metadatos de un documento electrónico (un texto, un sonido, una imagen digital, un programa de ordenador, etc. constituido por bits), difiera de los registros catalográficos tradicionales relativos a la información tangible (libros, revistas, vídeos, etc. constituidos por átomos). Esta disensión se debe a las siguientes particularidades de la nueva infraestructura de información, que ponen de manifiesto sus diferencias en cuanto a:

- La localización. Típicamente un registro de catálogo, tal como se presenta en un OPAC, hace referencia a una localización física asociada en la biblioteca. Un registro de metadatos hará referencia, sin embargo a localizaciones remotas, en muchas ocasiones no asociadas de ninguna manera con la institución que la realiza. Se requerirá pues, contemplar los detalles de los modos de acceso disponibles (ftp, http, etc.), así como las restricciones del mismo (*passwords*, etc.). Es frecuente además, que los objetos de información albergados en la Red se ubiquen en varias localizaciones de Internet y esta situación probablemente se multiplique ya que cada vez existen más *mirrors* o réplicas de páginas y sitios Web. Según esto, un registro de metadatos podría compararse con un catálogo colectivo desde el punto de vista que puede hacer referencia a varias localizaciones.
- El formato de los documentos. Otro de los problemas o características de los documentos electrónicos, susceptibles de representarse en un registro

¹⁰ Rachel Heery. *Review of Metadata...* *Op. cit.*, <http://www.ukoln.ac.uk/metadata/review.html>

¹¹ De acuerdo con los propios objetivos de esta investigación, es interesante que comparemos las características de los documentos de la Red, con los documentos tradicionales, para continuar con la metodología que hemos adoptado de tener como punto de referencia el mundo tradicional impreso, y para patentar aún más que existen contingencias entre un registro tradicional y un registro de metadatos, que nos hacen reafirmar que no son, en sentido estricto, lo mismo.

estructurado de metadatos, es el hecho de que un documento puede existir en diferentes formatos (html, pdf, PS, ASCII, etc.). En general, los distintos esquemas de metainformación permiten representar las versiones dispares de un mismo documento en un solo registro, mientras que en los catálogos tradicionales que describen documentos tangibles de información este hecho se contempla como distintas ediciones de la misma publicación, y se mantiene a un registro diferente para cada una de ellas.

- La falta de estabilidad¹². En muchas ocasiones los datos de la Red tienen poca estabilidad: los archivos se mueven, cambian de ubicación o simplemente desaparecen, de tal forma que su URL se vuelve inútil para la localización de

¹² Debido a la falta de estabilidad de la información de la Red, se están llevando a cabo diversas iniciativas para facilitar omnímodamente el acceso a los recursos Web. Estas iniciativas promovidas por el IETF (*Internet Engineering Task Force*) son: URN (*Uniform Resource Name/Number*) que funciona de forma similar al ISBN de los libros y que nombrará de manera unívoca a todos los objetos de Internet, y el URC (*Uniform Resource Characteristic*) que actuará como una citación del recurso. Sin embargo, ambas iniciativas van lentas, por ello existen otros mecanismos paralelos para asegurar que un recurso no se pierda si cambia de dirección. Uno de estos mecanismos es PURL (*Persistent Uniform Resource Locator*) <<http://purl.oclc.org>> que se desarrolla por OCLC para controlar los recursos de Internet en su catálogo colectivo. Vid. Rebecca Guenther. Naming Conventions for Digital Resources [documento HTML]. En: *MARC Homepage*. Washington, DC.: Library of Congress, 8 de julio de 1998. Disponible en: <http://cweb.loc.gov/marc/naming.html> (consultado el 24 de agosto de 1998).

Existen asimismo otros proyectos relacionados con la identificación unívoca y persistente de recursos Web, como por ejemplo: DOI (*Digital Object Identifier*) <<http://www.doi.org>> que es una norma ANSI/NISO Z39-84 desde el año 2000), WEBDAV (*World Wide Web Distributed Authoring and Versioning*) <<http://www.ics.uci.edu/pub/ietf/webdav>>, SICI (*Serial Item and Contribution Identifier*) <<http://sunsite.berkeley.edu/SICI>> también norma ANSI/NISO desde 1996, *Real Names* o *Handle System*, entre otros. Una descripción detallada del problema de la identificación de recursos Web así como de estas iniciativas, la encontramos en: Laurie Causton. *Identifying and Describing Web Resources* [documento PDF]. El Pub: Electronic Publishing, 17 de noviembre de 1998. Disponible en: http://www.elpub.org/assets/images/web_resources.pdf, p. 8-23 (consultado el 10 de octubre de 2000). La última norma en relación con la identificación de recursos Web, en el marco de la metainformación es: OpenURL <<http://www.sfxit.com/OpenURL/openurl.html>> que, desarrollado por la empresa Ex-Libris permite la transferencia normalizada de metadatos y/o identificadores de un objeto digital entre distintos servicios de información, a través de la Web. Tendremos oportunidad de retomar esta nueva tentativa al hablar de la interoperabilidad entre sistemas y de las bibliotecas híbridas.

un documento determinado. Por otra parte, los autores de las páginas cambian y actualizan los contenidos de las mismas aunque no varíen su URI, con lo cual un documento consultado en dos momentos distintos puede haber variado en forma y contenido de manera substancial a la primera vez que se consultó. Según esto, cabe preguntarse si los metadatos deben informar al usuario sobre el estado del documento en términos de calidad, o por el contrario tiene que ser una función del catalogador en el momento de la integración de tales registros en una colección digital concreta.

- Nivel de detalle (*granularity*). En los documentos tradicionales, el catalogador o indizador debe decidir el nivel de análisis, es decir, si el documento se va a describir como libro o por cada uno de sus capítulos, como un artículo o como una publicación periódica completa, etc., así como los campos que va a incluir en la descripción (autor, título, materias, etc.). Los registros MARC, en teoría, permiten relacionar niveles de análisis inferiores con el análisis del ítem completo. La posibilidad de describir el contenido de las publicaciones utilizando el índice o el resumen de las mismas ha existido siempre en el mundo bibliotecario tradicional pero no se ha utilizado mucho en los materiales monográficos, evidentemente por el coste de tiempo del catalogador. Con los recursos electrónicos se deben tomar soluciones similares, es decir, se debe decidir a qué nivel se desciende en la descripción. Por ejemplo si se trata de las páginas Web de un proyecto, se debe decidir igualmente qué ítems se describen: cada parte del proyecto, cada página, cada fichero, sus imágenes, o el sitio completo del proyecto.

El concepto de nivel de detalle en la descripción de recursos electrónicos llega más allá que el mero hecho de describir distintos niveles jerárquicos de las páginas. Teniendo en cuenta que los distintos formatos de metadatos se constituyen para describir DLOs, deben de ser flexibles para representar tanto

un ítem individual (v. gr. una página Web o una imagen), como recursos conceptualmente genéricos (una colección de páginas Web que forman un sitio o un conjunto de imágenes que forman una colección¹³).

- Naturaleza de la ubicación de los datos. Además de la dirección electrónica de un documento (URI/URL), es preciso consignar otras informaciones relativas al acceso, tales como las restricciones, direcciones de contacto, etc. Esta información no bibliográfica, no se puede comparar con los formatos de registro de publicaciones tradicionales, donde la ubicación es siempre física, concreta y no da lugar a más disquisición.

Aparte de señalar algunas de estas particularidades de los recursos electrónicos, Robin Wendler¹⁴, apunta otras características que hacen que estos sean difíciles de catalogar, procesar o describir:

- *La variabilidad en la presentación.* Dicho autor alude a las distintas configuraciones de los agentes de usuario (AU) que sirven como software cliente para visualizar las páginas Web. Esas diversas configuraciones locales o de estilo, determinan en muchos casos cómo aparece la información y por ende, según el navegador que se utilice, la apariencia y el orden de los contenidos, puede variar.
- *Formas no familiares de publicación.* Wendler hace una analogía entre las hojas sueltas y las publicaciones electrónicas, comentando al respecto que la

¹³ Esta idea de enfrentar el nivel de detalle (*granularity*, Vid. Glosario) de las descripciones que deben contemplar los distintos modelos de metadatos enfrentando el concepto de ítems *versus* colecciones, se ha discutido por ejemplo en torno al formato del Dublin Core, introduciendo en dicho formato la aplicación de calificadores a los elementos de relación. Cfr. 5.2.1.2.

¹⁴ Robin Wendler. Diversificación de actividades: habilidades y funciones catalográficas en la era digital. En: *Internet, metadatos y acceso a la información en bibliotecas y redes en la Era Electrónica*. Filiberto Felipe Martínez Arellano, Lina Escalona Ríos, comp. México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas; Infoconsultores, 2000, p. 37-38.

comunidad bibliotecaria nunca ha entendido muy bien qué hacer con las publicaciones en hojas sueltas. Lo más destacado de esta característica es que el contenido no está impreso y quizás aumente o se reemplace sin que, en muchas ocasiones haya una mención a que el estatus, la versión o la edición han cambiado. Esta característica tiene relación con la mutabilidad de este tipo de documentos, ya que salvo en el caso de los materiales electrónicos almacenados localmente en el lugar donde se realiza la descripción (catalogación convencional o asignación de metadatos) no hay garantía de que no se modifiquen radicalmente sin ninguna advertencia.

- *Relaciones poco definidas de los materiales electrónicos con otros formatos.* Un subconjunto de las publicaciones electrónicas se deriva, o está de alguna forma relacionado con las publicaciones de otros medios (normalmente el medio impreso pero también audio, vídeo, etc.). La dificultad aquí estriba en que no siempre es posible determinar el grado en que los DLOs cubren el contenido de esas variantes, ni tampoco es fácil precisar las diferencias funcionales entre los distintos formatos.
- Finalmente, Robin Wendler menciona la *ausencia de un receptor físico*. Los servicios de información fundados en objetos digitales carecen de punto de referencia claro y expreso a un usuario convencional tipo del que se conocen sus hábitos de consulta, etc. Se elimina pues, el punto clave (el usuario) sobre el que la mayoría de las bibliotecas han basado sus procedimientos de mantenimiento del catálogo y su política de acceso a la información.

Con todo, los recursos de información Web, los DLOs, objeto y fundamento de los metadatos, a diferencia de los documentos tradicionales, objeto de la catalogación, no son estables, tampoco responden a una selección previa apriorística, ni están organizados de una forma centralizada para su acceso por un usuario definido. A pesar de que muchos de los datos que requiere la información electrónica para diferenciar unívocamente unos recursos de otros pueden ser idénticos a los campos catalográficos tradicionales de las ISBD (autor, título, etc.), también es

preciso que los distintos modelos de metainformación tengan en cuenta estas connotaciones de disparidad y diversidad de los recursos de la Red; amén de que cada disciplina trate de plantear formatos adecuados a las características de su información particular. Existen pues, múltiples diferencias entre la actividad catalográfica y el registro de metadatos, fundamentalmente porque pertenecen a un paradigma de producción diferente; de igual forma que la relación que existe entre los metadatos y los objetos a los que se refieren, y los registros catalográficos y los documentos que referencian, es distinta. De esta forma, conceptos básicos de las reglas de catalogación como la fuente de información principal o las categorías de formatos como monografías, publicaciones seriadas, material cartográfico, vídeograbaciones, etc., se convierten en algo sin sentido, en un mundo en donde la visualización y el contenido de la información Web puede variar de un usuario a otro, sólo por el mero hecho de no tener instalada una aplicación concreta en su ordenador.

A pesar de estas diferencias y a pesar de que, desde nuestro punto de vista, los documentos digitales de la Web no deben catalogarse en sentido tradicional y estricto, es preciso reconocer que existen dos posturas a la hora de determinar el proceso técnico de los recursos electrónicos para optimizar su recuperación. Por un lado, la asignación de metadatos —postura que trataremos a lo largo de todo este capítulo, y que defendemos en esta tesis doctoral— y por otro, la catalogación de recursos electrónicos de manera convencional, como si se tratase de un tipo de documento más que integra la colección de la biblioteca. Esta doble visión sobre el procesamiento que deben tener los recursos electrónicos es la misma (ahora desde un punto de vista técnico-práctico) que veíamos en los planeamientos de Burnett, Bor Ng y Park¹⁵ en la doble aproximación teórica: la asignación de metadatos responde a la visión del concepto basada en la gestión de los datos electrónicos, y la catalogación, al enfoque del control bibliográfico.

¹⁵ Kathleen Burnett, Kwong Bor Ng, Syeon Park. *Op. cit.*, p. 1211-1212. *Cfr.* 3.2.

Antes de abordar la posibilidad del procesamiento convencional (catalogación) de los recursos electrónicos es necesario reflexionar ante varias cuestiones críticas: ¿merece la pena catalogar los recursos de la Red teniendo en cuenta la potencia de búsqueda de los nuevos sistemas de indización en texto libre?, ¿qué coste llevará aparejado?, ¿los formatos tradicionales de la catalogación —a saber, MARC/ISBD en sus múltiples versiones nacionales— son suficientes para la catalogación de documentos entendidos como objetos de información?, ¿cómo se debe afrontar la idiosincrasia de los recursos Web en cuanto a su localización, formato, falta de estabilidad, nivel de detalle, etc., a través de un proceso concebido para información tangible, como es la catalogación? y, sobre todo ¿qué catalogar?, es decir, la evaluación y selección de recursos Web para el tratamiento y la integración en sistemas de información. En este sentido Pedraza García¹⁶ apuntaba una serie de prioridades: *organizar y catalogar los recursos de Internet poseídos y mantenidos*

¹⁶ Manuel José Pedraza García. Aproximación a la catalogación de documentos electrónicos de acceso remoto. *Scire*, Enero-junio 1997, vol. 3, nº 1, p. 107. A pesar de consignamos estas prioridades básicas para la selección de documentos electrónicos, como hemos comentado, no profundizaremos en este tema, sin embargo queremos insistir una vez más en la importancia que tiene la calidad y la evaluación de recursos Web a la hora de decidir qué catalogar, a qué asignar metadatos o qué documentos incluir en una biblioteca digital formada por DLOs preexistentes (*subject gateways*). Por ello remitimos a las siguientes fuentes especializadas sobre el tema, *Vid.:*

- Lynn Elliott. *The ABC's of Evaluating WWW Sources* [documento HTML]. Panama City: Gulf Coast Community College Library, rev. 21 de febrero de 2001. Disponible en: <http://www.gc.cc.fl.us/library/researchhelp/webeval.htm> (consultado el 20 de mayo de 2001).
- Esther Grassian. *Thinking Critically About World Wide Web Resources* [documento HTML]. California: UCLA College Library, junio de 1995, rev. 6 de septiembre del 2000. Disponible en: <http://www.library.ucla.edu/libraries/college/help/critical/index.htm> (consultado el 3 de abril de 2001).
- Hope N. Tillman. *Evaluating Quality on the Net* [documento HTML]. Hopetillman.com, 1995, rev. 19 de marzo de 2001. Disponible en: <http://www.hopetillman.com/findqual.html> (consultado el 20 de mayo de 2001).
- Jan Alexander and Marsha Ann Tate. *Evaluating Web Resources* [documento HTML]. Wolfgram Memorial Library, Widener University, 8 de agosto de 1996, rev. 20 de abril del 2001. Disponible en: <http://muse.widener.edu/Wolfgram-Memorial-Library/webevaluation/webeval.htm> (consultado el 20 de mayo de 2001). Se trata de un módulo para el aprendizaje de la evaluación de recursos Web que incluye presentaciones, ejemplos, listas de comprobación para chequear la calidad de distintos recursos desde un punto de vista temático y sobre todo, consigna también una amplia bibliografía complementaria sobre el tema.

por el sistema local¹⁷; los materiales de investigación significativos que tienen interés para el usuario local; herramientas significativas que puedan mejorar los servicios de referencia y herramientas significativas que pueden actualizar el conocimiento y mejorar las habilidades del equipo técnico de la biblioteca. De alguna manera estas prioridades o esta selección tácita de recursos de Internet para su tratamiento la veremos en los proyectos de catalogación *strictu sensu* de la Red, de los cuales hablaremos a continuación.

La idea de catalogar los recursos de Internet o quizás mejor, la actitud de por qué no catalogar los recursos de Internet, surge en el seno de OCLC a principios de los años 90, y comenzaría con la catalogación piloto de 300 recursos de Internet por 30 catalogadores con experiencia probada. Posteriormente, en 1995, la publicación de *Cataloging Internet Resources: A Manual and Practical Guide* por parte de la reconocida catalogadora de materiales audiovisuales Nancy Olson¹⁸, así como la adopción del campo 856 (localización electrónica y acceso) en el formato USMARC, que permitía vincular el registro bibliográfico con el recurso electrónico remoto al

¹⁷ En este aspecto también coincide Weber que opina que las bibliotecas pueden optar, por ejemplo por catalogar todos los recursos que ha adquirido o suscrito o los listados en la página Web de la biblioteca [...] Mary Beth Weber. Factors to be Considered in the Selection and Cataloging of Internet Resources. *Library Hi Tech*, 1999, vol. 17, n° 3, p. 300. Añade además una idea difundida por la *University of Rochester River Campus Libraries* en su guía para la catalogación de recursos electrónicos publicada en 1997 <<http://www.lib.rochester.edu/cat/catguide.htm>>: *El catálogo de la biblioteca ha excedido su papel tradicional. Ya no es un inventario de lo que la biblioteca posee sino más bien de a lo que la biblioteca tiene acceso.*

¹⁸ Nancy B. Olson. *Op. cit.*, <http://www.oclc.org/oclc/man/9256cat/toc.html> A pesar de que la trayectoria profesional de Olson hizo que este manual fuese un punto de referencia sobre el tema, otros autores difieren de sus planteamientos. Por ejemplo. Hsieh-Yee propone lo que ella denomina *augmented minimal level cataloging standard*, un estándar ampliado para la catalogación a un nivel mínimo de los recursos electrónicos. Además esta última autora piensa que, los registros creados en el seno de OCLC gracias a la experiencia de sus catalogadores y a su infraestructura, deben competir con otros sistemas de búsqueda en Internet. Ingrid Hsieh-Yee. *Modifying Cataloging Practice and OCLC Infrastructure for Effective Organization of Internet Resources* [documento HTML]. En: *OCLC Internet Cataloging Colloquium (1996. San Antonio)*. OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/hsieh.htm> (consultado el 19 de abril de 2001).

que representaba, establecerían las coordenadas básicas para que OCLC desarrollara su proyecto de catálogo InterCat¹⁹. Uno de los objetivos de este proyecto fue, entre otros, evaluar la eficacia de la utilización del formato USMARC, incluyendo la localización electrónica y la información sobre el acceso (856), para proporcionar la descripción de los objetos de información disponibles de forma remota en Internet²⁰.

¹⁹ *Building a Catalog of Internet Resources* (nombre formal que recibió InterCat) fue un proyecto muy ambicioso en el que se invirtieron 62.000\$ del Departamento de Educación de Estados Unidos desde el 1 de octubre de 1994 hasta el 31 de marzo de 1996. Trataba de coordinar el esfuerzo de bibliotecas y centros de información de todo el mundo para crear, implementar y evaluar una base de datos de registros bibliográficos en formato USMARC donde apareciese, perfectamente completada, la información relativa a la localización y acceso de los recursos disponibles en Internet. El catálogo fruto del proyecto InterCat estuvo accesible en el URL: <http://purl.org/net/intercat>, pero actualmente no se puede consultar, lo que corrobora, en cierto sentido, la validez de este proyecto como prototipo de aplicación del campo 856, pero su escasa vigencia como método de organización y recuperación de recursos electrónicos. El total de registros en InterCat hasta el 15 de julio de 1998 (fecha en que por última vez, pudimos acceder a este catálogo) era de 40.192: muchos, teniendo en cuenta lo costoso de la catalogación en MARC, pero poquísimos, teniendo en cuenta la "infinitud" o inabarcabilidad de recursos de información o documentos entendidos como objetos, disponibles en la Red.

²⁰ Para una visión pormenorizada del funcionamiento del campo 856 del formato MARC 21 tras la revisión en la conferencia de la ALA de junio de 1999, donde se detallan sus diversos indicadores y códigos de subcampo, se recomienda la consulta de la Guía de uso que ha elaborado la Library of Congress sobre este campo. *Guidelines for the Use of Field 856* [documento HTML]. Washington, D. C.: Library of Congress, Network Development and MARC Standards Office, rev. 9 de agosto de 1999. Disponible en: <http://www.loc.gov/marc/856guide.html> (consultado el 23 de abril de 2001). El campo 856 se puede utilizar, en esta misma línea del control bibliográfico, como una forma de crear registros de autoridad, como se puede observar en este ejemplo tomado de: John Riemer. Adding 856 Fields to Authority Records: Rationale and Implications. *Cataloging & Classification Quarterly*, 1998, vol. 26, n° 2, p. 7:

100 10 Van Buren, Abigail, \$d 1918- [...] 856 4_ \$u <http://www.uexpress.com/ups/abby>

Además de este campo específico de localización y acceso, la catalogación convencional en MARC también requiere cambios en la cabecera del registro. *Vid. Guidelines for Coding Electronic Resources in Leader/06* [documento HTML]. Washington, D. C.: Library of Congress, Network Development and MARC Standards Office, rev. 7 de enero de 1999. Disponible en: <http://lcweb.loc.gov/marc/ldr06guide.html> (consultado el 23 de abril de 2001).

Además de este proyecto de OCLC se llevaron a cabo otros paralelos, liderados por otras instituciones en el ámbito europeo, como por ejemplo CATRIONA²¹ (*CATaloguing and Retrieval of Information Over Networks Applications*). En el caso de CATRIONA, a pesar de estar basado también en la catalogación de recursos electrónicos y en el campo 856 del USMARC, la idea subyacente iba más allá que la mera catalogación cooperativa, se trataba de crear un catálogo distribuido de recursos de Internet gracias al estándar Z39.50 y a través de Interfaces Gráficas de Usuario (GUI) tipo OPAC. El cliente se denominaba GeoPac, y podía buscar en OPACs Z39.50 remotos recuperando registros MARC con el URL en 856 \$u. CATRIONA trataba de proporcionar a los usuarios de biblioteca una forma confiable de recuperar información, por un lado, utilizando los estándares bibliotecarios tradicionales para la descripción y recuperación y por otro, entendiendo los documentos electrónicos como objetos, lo que en este proyecto denominaban, como ya hemos mencionado, EIOs, tratando de integrar el acceso con otros recursos no electrónicos. En CATRIONA ya existía la certeza de que crear un gran OPAC para los recursos de Internet no era posible por múltiples razones, entre otras, cuestiones económicas y diferencias internacionales o de idioma.

Al igual que ocurre con el proyecto InterCat, el catálogo o sistema de acceso, en este caso distribuido, de CATRIONA ya no está disponible en Internet. Sin embargo, tuvo una segunda fase en el seno de eLib, de junio de 1996 a diciembre de 1998, que pretendía inicialmente *analizar las estrategias de creación y gestión de recursos electrónicos para la investigación y la docencia en las universidades escocesas*²². CATRIONA II incluía dentro de sus objetivos continuar con la misma

²¹ Para más información sobre este proyecto es interesante leer el *estudio de factibilidad* que realizaron miembros del departamento de Investigación y Desarrollo de la British Library en 1995. Dennis Nicholson, et al. *Cataloguing the Internet... Op. cit.*, <http://bubl.ac.uk/org/catriona/cat1rep.htm>

²² Dennis Nicholson, Martin Smith. *Electronic Resource Creation and Management at Scottish Universities: Survey Results and Demonstrator Service Progress* [documento HTML]. *Ariadne*, 26- [cont.]

idea de crear una interfaz Z39.50 para recuperar información electrónica para la educación a distancia y hoy por hoy está vinculado a un proyecto mayor de campus virtual en las universidades de Strathclyde <<http://cvu.strath.ac.uk/library>> y Napier <<http://catriona.napier.ac.uk>>, que aprovechan el modelo distribuido de recuperación de información difundido en CATRIONA I, pero que cada vez más, incluyen la organización de los recursos electrónicos a través de metadatos²³.

Los proyectos como InterCat o la primera parte de CATRIONA sirvieron para crear conciencia a los bibliotecarios y catalogadores sobre la organización de recursos de Internet y el formato MARC, así como para incentivar la necesidad de nuevas alternativas a la catalogación.

A raíz de proyectos como estos empezaron a surgir talleres, encuentros y gran cantidad de literatura que discutía, en mayor o menor medida, la viabilidad del MARC para describir recursos electrónicos, e incluso, como modelo de metainformación²⁴. Dentro de la vertiente estrictamente catalográfica abocada al

June-1998, issue 14,. Disponible en: <http://www.ariadne.ac.uk/issue14/catriona> (consultado el 22 de abril de 2001).

²³ La orientación del proyecto CATRIONA II, aunque en esta ocasión lo citamos por su vinculación en la evolución de la descripción de recursos electrónicos a través del MARC, corrobora la tesis que defendemos aquí, la aplicabilidad de metadatos en contextos finitos de información, puesto que las universidades escocesas están estudiando e incorporando metadatos a sus proyectos de biblioteca digital como soporte a la educación a distancia. Así lo demuestra, por ejemplo, la *Clyde Virtual University Library* <<http://cvu.strath.ac.uk/library>> cada vez más vinculada a los modelos DC e IMS (*Instructional Management Systems*); o el proyecto STAR (*Strathclyde Authored Resources*) que recomienda la asignación de metadatos a los autores que crean recursos electrónicos para la docencia y la investigación en dicha Universidad.

²⁴ Uno de los eventos pioneros en este sentido fue el *Seminar on Cataloging Digital Documents* <<http://marvel.loc.gov/catdir/semdigdocs/seminar.html>> coorganizado por la Universidad de Virginia y la Biblioteca del Congreso de los Estados Unidos en octubre de 1994. Posteriormente, surgido justamente a raíz del proyecto InterCat y en el seno de OCLC, debemos destacar también el *OCLC Internet Cataloging Colloquium* <<http://www.oclc.org/oclc/man/colloq/toc.htm>> celebrado en San Antonio, en enero de 1996, al que ya hemos hecho alusión. Sobre la aprehensión del formato MARC como modelo de metadatos profundizaremos en el capítulo 6 de esta tesis doctoral al hablar de la normalización. *Cfr.* Capítulo 6, nota 20.

control bibliográfico tradicional, Amanda Xu²⁵, por ejemplo, defiende la organización de recursos electrónicos a través de formatos MARC y de las AACR2, basándose en los beneficios que pueden reportar estos estándares para la recuperación de información. Para defender esta postura se basa, en primer lugar, en que estos formatos han sido seleccionados previamente por bibliotecarios para satisfacer las necesidades de una comunidad determinada. Por otra parte, la catalogación tradicional está controlada, de manera concreta y formal, a través de la descripción bibliográfica, del control de autoridades y del análisis de materias. Esta autora alude también en su planteamiento, a los sistemas de automatización de bibliotecas, ya que estos se fundamentan en el formato MARC y en los estándares habituales de la catalogación, tradicionalmente utilizados y probados, que además se han desarrollado para manejar grandes cantidades de datos y ejecutar búsquedas sofisticadas. Finalmente señala la ventaja de la integración en el sentido de que a través de la utilización del formato MARC, los recursos de Internet pueden integrarse con otros registros bibliográficos que ya han sido organizados en función de dicho formato. Desde nuestro punto de vista, la integración de recursos electrónicos en un catálogo único es la razón exclusiva que podríamos aceptar para la catalogación tradicional de los objetos de información Web, ya que entendemos que, en determinados contextos, puede resultar de interés tener un acceso idéntico en un OPAC tanto a recursos de la Red como a recursos de información tangible (libros, revistas, mapas, CDs, disquetes, etc.) que se hayan catalogado previamente y que formen parte de la colección de una unidad de información determinada. Sin embargo, como veremos a lo largo de esta investigación, el futuro de la compatibilidad en la recuperación en entornos híbridos de información tenderá más a la interoperabilidad entre el MARC y

²⁵ Amanda Xu. Accessing Information on the Internet: Feasibility Study of USMARC Format and AACR2 [documento HTML]. En: *OCLC Internet Cataloging Colloquium (1996. San Antonio)*. OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/xu.htm> (consultado el 19 de abril de 2000).

los nuevos modelos de codificación de la información, que a la imposición categórica del MARC para procesar recursos electrónicos²⁶.

Eric Jul también defiende la tendencia catalográfica tradicional de los recursos electrónicos apoyándose en su caso, en que los catalogadores suponen una mano de obra especializada, junto a los estándares comunes, prácticas y sistemas que pueden aplicarse para mejorar la descripción y el acceso a Internet donde los metadatos no son más que una precatalogación en la publicación²⁷. Utiliza igualmente el argumento de la integración de las bases de datos bibliográficas que incluyan en los OPACs locales tanto recursos librarios como electrónicos, y además opina que las bibliotecas que enfrenten el trabajo de la catalogación de recursos de Internet ganarán en la comprensión que, a través de la experiencia, ayudará a moldear la evolución futura de Internet. Este autor introduce otro factor que debemos considerar: el coste. Jul hace referencia a que el coste de la catalogación de recursos electrónicos se atribuye principalmente a que *es un trabajo intensivo que requiere tareas especializadas*²⁸. Desde nuestro punto de vista el coste de "catalogar la red" (ya sea de forma centralizada o distribuida) es demasiado alto, en términos de tiempo para la rentabilidad que puede producir a efectos de recuperación generalizada en Internet, y la asignación de metadatos debe de ser una tarea sustitutiva de la catalogación no un trabajo previo y acumulable, en dos momentos distintos, al esfuerzo de la catalogación convencional.

²⁶ Cfr. 8.3.2.

²⁷ Erik Jul. Cataloging Internet Resources: Survey and Prospectus. *Bulletin of the American Society for Information Science*, October/November 1999, vol. 24, nº 1, p. 7-8. Lo más curioso del planteamiento de este autor es que piensa los metadatos no son una alternativa a la catalogación tradicional, sino un soporte a ésta. Dice expresamente (p. 8) que los metadatos utilizados por personas o profesionales no bibliotecarios (*others*) pueden ser poco más que un "revoltijo de campos incontrolados" (*...used by others, metadata may be little more than a mishmash of uncontrolled fields*).

²⁸ *Ibid.*

Siendo el formato MARC uno de los baluartes en la defensa de la catalogación en sentido estricto de documentos electrónicos, y pudiendo considerarlo *latu sensu*, un modelo más de metainformación, es el momento de pronunciarnos en cuanto a su uso como estándar para la descripción de recursos electrónicos. Son muchas las razones que se podrían esgrimir al respecto, teniendo en cuenta todo lo que hemos dicho hasta ahora:

- Casi todos los autores, sobre todo los que hemos señalado dentro de la corriente del control bibliográfico, reconocen al formato MARC como un modelo más de metadatos, incluso dentro del tipo más complejo y estructurado (banda 3)²⁹.
- Se trata de un conjunto de formatos consistentes con la norma ISO 2709 que tienen gran solidez y aplicación como estándares internacionales y que, en consonancia con el CBU, han dirigido la creación, distribución e intercambio de información bibliográfica durante los últimos cuarenta años³⁰.
- A pesar de la demora, la comunidad MARC ha manifestado múltiples procedimientos para actualizar este formato y adaptarlo a la descripción y gestión de documentos electrónicos. Por ejemplo, en 1998 se creó una Definición de Tipo de Documento (DTD) en SGML³¹, posteriormente a partir

²⁹ Vid. Capítulo 3, Tablas 8 y 9.

³⁰ No obstante hay que tener en cuenta, como afirma Lluís Anglada, que *después de treinta años de su utilización y en informática, no hay desarrollo que pueda mantener ninguna lozanía tecnológica tanto tiempo* Lluís M^a Anglada i Ferrer. Biblioteca Digital ¿mejor, peor o sólo distinto? *Anales de Documentación*, 2000, n^o 3, p. 34. Retomaremos la valoración del formato MARC como norma o estándar de intercambio bibliográfico en el apartado 6.1. de este trabajo.

³¹ *MARC DTDs Document Type Definitions: Backgrounds and Development* [documento HTML]. Washington, D.C.: Library of Congress, Network Development and MAR Standards Office, 22 de mayo de 1998. Disponible en: <http://www.loc.gov/marc/marcdtd/marcdtdback.html> (consultado el 13 de noviembre de 2000).

de MARC21 se ha ido desarrollando un conjunto de procedimientos y de software para la interoperabilidad entre el MARC y XML³².

- Más de cuarenta años de uso del formato MARC implican una gran inversión en software y aprendizaje por parte de las bibliotecas y de los profesionales de la información.
- El uso de MARC para la descripción de recursos electrónicos permite la integración de estos en un catálogo único de biblioteca³³.

A pesar de todas estas afirmaciones a favor, vemos diversas desventajas a la utilización del MARC como una estructura de metadatos, entre las que es indispensable destacar:

- Lo costoso de la construcción y mantenimiento de registros en formato MARC.

³² Para más información, consultar el sitio oficial del formato MARC 21 en la *Library of Congress. MARC Development* <<http://lcweb.loc.gov/marc/development.html>> y *MARC, SGML and XML* <<http://lcweb.loc.gov/marc/marcsgml.html>> así como las herramientas para pasar del formato MARC a otros modelos de estructuración de la información <<http://lcweb.loc.gov/marc/marctools.html#web>>. También son interesantes las comunicaciones presentadas en la *International Conference on The Future of Communication Formats (1996. Ottawa)* [documento HTML]. Hull (Canadá): Banque International d'information sur les États francophones, rev. 23 de febrero de 1996. Disponible en: <http://www.acctbief.org/avenir/proceed.htm> (consultado el 10 de abril de 2001), algunas de las cuales ya hemos citado (*Vid. suprà*. Nota. 8). En este congreso, organizado por la Biblioteca Nacional de Canadá en octubre del 96, se habló fundamentalmente de los entonces "nuevos" formatos de descripción bibliográfica, especialmente SGML y las expectativas del formato MARC ante los documentos electrónicos, incluyendo, aunque tímidamente todavía, temas como Z39.50 y metadatos.

³³ Así se dice explícitamente en el proyecto CATRIONA: *la necesidad de integrar la recuperación de Objetos de Información Electrónica (EIOs) con la recuperación de recursos de información no electrónicos implica que es importante elegir un estándar bibliotecario aceptado para la descripción de estos recursos, como MARC*. Dennis Nicholson, et al. *Cataloguing the Internet... Op. cit.*, <http://bubl.ac.uk/org/catriona/cat1rep.htm>

- La complejidad implícita de este formato requiere personal cualificado así como un software específico.
- Fundamentalmente, a pesar de su evolución, de la inclusión del campo 856, de las conversiones a DTDs en SGML o XML, y de sus múltiples defensores como modelo de metadatos para los recursos electrónicos, el formato MARC se constituyó sobre el modelo de las fichas de catálogo y está eminentemente orientado al libro impreso y a los soportes tangibles de información.
- Además, el MARC no contempla la inserción de metadatos estructurales ni administrativos, ni tiene el nivel de adaptabilidad y utilidad de los nuevos esquemas de metainformación.

Quizás por el reconocimiento de estas fortalezas y debilidades del MARC surge el proyecto CORC (*Cooperative Online Resource Catalog*), una vez más en el seno de OCLC. CORC utiliza tanto el MARC como el Dublin Core para crear una base de datos de recursos "de calidad" de Internet con perspectivas de ser el *WorldCat para los recursos electrónicos, o en el mejor de los casos, podría completar los servicios de búsqueda existentes, convirtiéndose en el motor de búsqueda de facto en el ámbito académico*³⁴. Al igual que en otros proyectos previos de catalogación, utiliza la cooperación para construir sus colecciones. En un principio comenzó con los

³⁴ Norm Medeiros. Making Room for MARC in a Dublin Core World. *Online*, 1999, vol. 23, n° 6, p. 60. Ante la referencia que hace Medeiros al *WorldCat* debemos recordar que *WorldCat* es el catálogo unificado de OCLC <<http://www.oclc.org/oclc/menu/colpro.htm>> que el 20 de septiembre de 2001 introdujo el registro n° 48 millones, y que representa 400 lenguas.

Al fin y al cabo CORC no es más que el traspaso de la filosofía del modelo cooperativo de *WorldCat* a los recursos electrónicos. A fecha de 21/11/2000 CORC contaba con 384.942 registros y el 11/12/2001, 560.085 registros, lo que implica aproximadamente un crecimiento de 13.500 registros al mes, en el último año.

registros del proyecto InterCat y de NetFirst³⁵, pero actualmente depende del trabajo de las bibliotecas participantes³⁶.

Una de las principales características de CORC es la interoperabilidad: los participantes en el proyecto pueden incluir registros tanto en MARC como siguiendo los elementos del modelo DC; éstos se almacenan en XML y pueden disponerse o exportarse también en formato Dublin Core (con sintaxis XML/RDF o HTML) o en MARC21 en los sistemas locales de las bibliotecas. Lo más interesante de CORC, desde nuestro punto de vista, es que es un legado del estado actual de los metadatos y del control y descripción de DLOs desde una perspectiva eminentemente bibliotecaria. Supone, por un lado, la asunción de los metadatos como estándares para la recuperación de información de calidad, y por otro, la revalorización de las técnicas bibliotecarias de control y acceso por materias, a través de la generación de *pathfinders*³⁷ o selecciones automáticas de recursos sobre un tema. Además CORC supone el reconocimiento de la necesidad de crear sistemas que interoperen entre

³⁵ NetFirst es una base de datos Web accesible sólo para los clientes del servicio de bases de datos bibliográficas de OCLC FirstSearch. Se trata de una base de datos "de autoridad" creada por el personal de OCLC donde se presenta la indización y el resumen de los registros incluyendo los encabezamientos de materia de la LC y los códigos de clasificación de la DDC. Para más información sobre NetFirst, consultar: <http://www.oclc.org/oclc/netfirst> y sobre el servicio FirstSearch: <http://www.oclc.org/firstsearch>; http://www.oclc.org/oclc/fs_spanish/frames_man.htm (en español).

³⁶ A fecha de 14/11/2000 participaban en el proyecto CORC casi 600 bibliotecas de todo el mundo, de las cuales sólo tres son españolas: el CINDOC, la biblioteca de la Universidad de La Rioja y la de la Universidad Politécnica de Valencia. Con posterioridad a esa fecha, se ha incorporado también al proyecto la biblioteca de la Fundación Juan March <<http://www.march.es>>. Para más información sobre las instituciones participantes, consultar: <http://www.oclc.org/corc/about/founding.shtm>

Isabel Zulbedia, responsable de referencia e información del área de Documentación Científica de la Biblioteca de la Universidad Politécnica de Valencia, describe el proceso de adhesión al proyecto CORC de esta institución española, antes de que fuese un servicio de valor añadido de OCLC (30 de junio de 2000) y fuese necesario subscribirse para participar. M^a Isabel Zubeldia Lauzurica. Cooperación frente al caos en Internet: CORC, una propuesta de trabajo. En: *Jornadas Españolas de Documentación* (7. 2000. Bilbao). *Op. cit.*, p. 597-606.

³⁷ *Vid.* Glosario.

distintos formatos y aún tres factores clave para el desarrollo del procesamiento de los recursos electrónicos: tecnología, cooperación y gestión bibliotecaria.

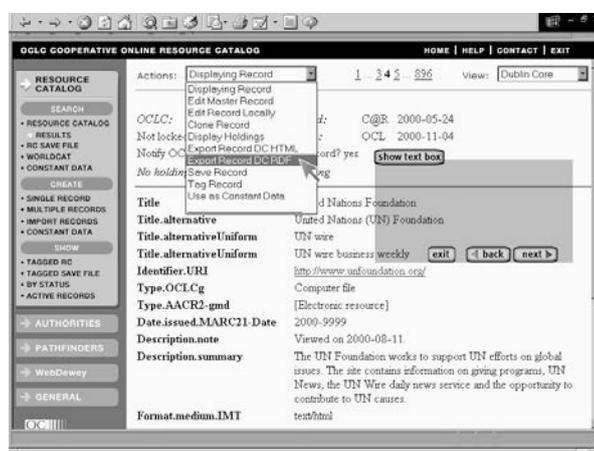


Fig. 2. CORC. Visualización DC de un registro, exportable a RDF³⁸

En este punto que, a pesar de que traemos aquí el planteamiento convencional de la catalogación como una forma de procesar, describir u organizar los recursos electrónicos, y que los autores y proyectos que citamos forman parte del acervo de reflexiones en torno a este tema, debemos reconocer que estas tentativas formaban parte de los proyectos y documentos que los atestiguan, surgidos en torno al periodo

³⁸ Esta imagen pertenece a la Demo interactiva de CORC <<http://www.oclc.org/corc/learning/demo>> (Requiere la instalación del *plug-in* Macromedia Shockwave Player <<http://www.macromedia.com/shockwave/download/>>).

Lo que aparece en la Fig. 2 es el resultado final de la base de datos de CORC, sin embargo detrás de este producto están varios años de investigación llevados a cabo en el seno de OCLC como MANTIS o Scorpion. Mantis está constituido por un conjunto de herramientas basadas en RDF para crear sistemas de catalogación Web con definiciones arbitrarias de metadatos y diferentes interfaces; Scorpion parte del reconocimiento de que la información por materias es la clave de la recuperación de información avanzada y de la navegación, por ello ha creado distintas herramientas para el reconocimiento automático de materias basándose en esquemas (*schemes*) bien conocidos en el ámbito bibliotecario como el Sistema Decimal de Dewey (DDC). Para más información sobre la tecnología desarrollada para soportar CORC. Vid. Keith Shafer. Scorpion Helps Catalog the Web. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, nº 1, p. 28-29 y Keith Shafer. *Mantis, A Flexible Cataloging Toolkit* [documento HTML]. OCLC, 22 de octubre de 1998. Disponible en: <http://orc.rsch.oclc.org:6464/toolkit.html> (consultado el 20 de septiembre de 1999).

temporal de 1994 a 1997. En estos "primeros años" sólo se dedicaba a los metadatos un pequeño apartado bajo marbetes como *métodos alternativos de la catalogación de documentos digitales*³⁹, *mirando al futuro*⁴⁰ o simplemente *un enfoque* más⁴¹ dentro de las propuestas para el tratamiento de la información de la WWW, y en todos ellos se reconocía el estado incipiente de los nuevos formatos de metainformación.

Tanto las teorías como los proyectos en curso evidencian que existen dos posibilidades de tratamiento técnico de los recursos electrónicos: la catalogación tradicional y la asignación de metadatos, aunque ambas tendencias tienden a concurrir como demuestran iniciativas como CORC. De igual forma que al hablar del concepto nos adscribíamos a la vertiente estructuralista y a la teoría de los metadatos en sentido estricto, a la hora de hablar de la visión práctica, esto es, de su tratamiento real en el seno de un sistema de información de Internet, defendemos la postura de la creación de metadatos, cualquiera que sea la forma en que estos se asignen (que discutiremos en el apartado siguiente), y el formato o esquema que se adopte (que trataremos en el capítulo 5) tomando aquellos aspectos de la catalogación tradicional que puedan coadyuvar a una mejor gestión de la metainformación y por ende, de la recuperación de información.

Desde nuestro punto de vista, se requiere un proceso de adaptación intelectual para aplicar los modelos abstractos, las reglas, las prácticas, los códigos y los procedimientos de la Biblioteconomía y Documentación a las características de las publicaciones electrónicas. Los metadatos cubren el vacío entre los dos extremos del tratamiento y recuperación de información en Internet: el de la catalogación minuciosa, completa y tradicional de recursos electrónicos y su acceso a través de

³⁹ *Alternative methods of cataloging digital documents* (1997). Kyle Banerjee. *Op. cit.*, p. 16-18.

⁴⁰ *Looking ahead* (1997). Erik Jul. *Op. cit.*, p. 8.

⁴¹ *El enfoque metadata* (1997). Manuel José Pedraza García. Aproximación a... *Op. cit.*, p. 103-105.

motores de búsqueda⁴². Esa adaptación intelectual a la que hacemos referencia, no implica que se deban llevar a cabo, en general, todas las actividades al nuevo entorno, y como dice Brisson⁴³, la catalogación de los materiales electrónicos para los OPACs es un ejemplo relevante de la cautela que hay que tener con la simple aceptación de las actividades y prácticas bibliotecarias. Creemos que se trata, más que de una aceptación del modelo de gestión de información que implican los metadatos, de una adaptación. El planteamiento que subyace en los sistemas de metadatos supone la adaptación, adecuada al nuevo medio electrónico, de la descripción y organización de recursos para su localización y recuperación.

4.2. Quién crea los metadatos

Una vez diferenciada la catalogación frente a la asignación de metadatos que, como vimos, son soluciones distintas y alternativas, inclusive suplementarias para algunos autores, a la hora de describir los documentos entendidos como objetos de información en la Red; y decantada nuestra postura de aplicar a nuevos contextos informativos (DLOs), nuevas soluciones para la organización y recuperación (metadatos), corresponde abordar ahora la cuestión fundamental de determinar quién es el agente que debe crear esos metadatos y dónde y cómo almacenarlos. Aunque en este apartado apuntaremos reflexiones de carácter general, desde nuestro punto de

⁴² Sobre la catalogación convencional a través del MARC, hemos hablado ampliamente a lo largo de este apartado; sobre la recuperación de recursos Web a través de motores de búsqueda profundizaremos en el capítulo 7. La idea de que la gestión de metadatos está entre el tratamiento escrupuloso de la catalogación y la ausencia de tratamiento o automatización de la indización de los motores de búsqueda, se puede colegir de las ideas de autores australianos como Talmacs. Kerrie Talmacs. Scott and the Race for the South Pole: the Triumph of Metadata. En: *National Cataloguing Conference (13. 1999. Brisbane): Charting the Information Universe* [documento HTML]. Sydney: University of New South Wales Library, 14 de octubre de 1999, rev. 1 de noviembre de 1999. Disponible en: <http://www.library.unsw.edu.au/~eirg/conf/natcat99.html> (consultado el 26 de julio de 2000).

⁴³ Roger Brisson. *Op. cit.*, p. 29.

vista hay que diferenciar dos contextos, semejantes, no excluyentes e incluso complementarios, pero en definitiva distintos:

- Los metadatos creados con la finalidad de mejorar la recuperación genérica en la Web a través de motores de búsqueda, donde se necesita adoptar un paradigma de metadatos que instigue a los autores a describir sus propios recursos para que estos tengan una recuperación más eficaz.
- La metainformación asignada en un contexto informativo concreto o biblioteca digital, donde los requisitos descriptivos y estructurales deben responder a un canon de precisión mayor y determinado *a priori* según las características de los DLOs que constituyan dicho sistema de información electrónico.

Inicialmente partimos de que la función del creador de metadatos debe de ser distinta de la del bibliotecario-catalogador tradicional, tanto desde el punto de vista de su formación, perfil y conocimientos, como de su propia filosofía de trabajo. En este contexto de la información electrónica que estamos describiendo, Butterfield⁴⁴ ve al "catalogador", al nuevo catalogador⁴⁵, por un lado como *un mediador* y por otro

⁴⁴ Kevin L. Butterfield. *Cataloger's and the Creation of Metadata Systems: A Collaborative Vision at the University of Michigan* [documento HTML]. En: *OCLC Internet Cataloging Colloquium (1996. San Antonio)*. OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/butter.htm> (consultado el 19 de abril de 2000). Butterfield utiliza para definir al nuevo catalogador como mediador-creador, la experiencia surgida en la Universidad de Michigan que adoptó las cabeceras TEI (*Vid.* 5.2.2.1.), donde los catalogadores, a la par que desarrollaron destrezas en UNIX, SGML, etc. mediaron en el desarrollo de etiquetas de contenido, coadyuvando a la creación de unas cabeceras de metadatos más descriptivas y normalizadas.

⁴⁵ Aunque nos hemos referido ya al "catalogador" (entre comillas) de los recursos electrónicos, en coherencia con lo que hemos afirmado que, los metadatos no son catalogación en sentido estricto, igualmente la persona que asigna los metadatos no es un catalogador. Entre las denominaciones más acertadas para designar al nuevo agente que creará la metainformación a los recursos electrónicos se encuentra la de *creador de metadatos (metadata creator)*; no obstante, esta nueva figura recibirá otras denominaciones en contextos concretos, como p. ej.: *autores, gestores de depósitos documentales*, etc. cuando la labor de creación de los metadatos recae en la persona/s que desempeñan estas funciones.

como *un creador*. Los catalogadores proporcionan una fuerte influencia de la normalización en los nuevos sistemas de descripción y acceso a través de metadatos, actúan pues como intercesores entre las reglas de catalogación y las nuevas formas de describir y/o estructurar la información. Por otra parte, el catalogador es un participante activo en la creación de sistemas de metadatos. De hecho múltiples de los proyectos de desarrollo de esquemas de metainformación parten de la comunidad bibliotecaria y concretamente de los catalogadores —esta evidencia es mayor en el caso del Dublin Core—, que definen y proponen etiquetas y componentes de los modelos de metainformación. Sin embargo, en nuestra opinión, cada vez más, los modelos de edición que incorporan el acceso electrónico imponen la asignación de metainformación vinculada al proceso de creación del documento.

Sobre esta cuestión hay un consenso más o menos generalizado acerca de que *el propio texto necesita tener una información estandarizada sobre sí mismo*⁴⁶, en el sentido de que debe contener "datos sobre datos" de una manera estructurada o normalizada. Incluso los sectores más axiomáticos en la defensa de la catalogación de todos aquellos recursos eventualmente interesantes para una unidad de información, independientemente de que éstos estén en formato impreso, digital u otro, advierten esta necesidad. Pat Oddy⁴⁷ opina que el catalogador puede usar esa información normalizada que debe contener el propio documento, para construir el registro bibliográfico o que esta información puede constituir incluso, un registro de catálogo en sentido estricto, de igual forma que en el mundo impreso algunos libros llevan los datos catalográficos en la propia publicación (*Cataloging In*

⁴⁶ Pat Oddy. *Op. cit.*, p. 167. Oddy, menos taxativa que Jul (*Cfr. supr.* Nota 27), se refiere a que un registro de metadatos puede formar un registro en sí mismo de un Sistema de Recuperación de Información en Internet (SRII) o bien puede servir de información para crear un registro de un catálogo accesible vía Web (OPAC Web o WebPAC), o un registro concreto, por ejemplo del proyecto InterCat.

⁴⁷ Pat Oddy. *Ibid.*

*Publication*⁴⁸). La teoría de la catalogación en fuente (CIP) implica la descripción en el momento más cercano posible a la producción del documento. En el caso de la documentación electrónica esto consiste en incluir o, utilizando la terminología más difundida en el contexto de los metadatos, embeber/incrustar⁴⁹ en el código de un documento en texto completo, una cabecera estructurada (por ejemplo y de forma más sencilla, dentro de la etiqueta <HEAD> en HTML) donde se registra la información descriptiva del documento, de tal manera que tengamos, en una misma unidad lógica, tanto el documento como su representación informativa (metadatos).

De alguna forma, el *cuasi-consenso* del ECIP (*Electronic Cataloging in Publication*), mitiga el problema del ¿cuándo?, al reconocer, la mayoría de los autores, que los metadatos deben establecerse lo más cerca posible del momento cero o momento de creación del documento⁵⁰; aunque también hay quienes advierten los riesgos que implica la introducción de metainformación directamente durante la fase de edición de los datos o inmediatamente antes de su publicación en el servidor. Por

⁴⁸ Típicamente el programa *Cataloging in Publication* de la Biblioteca del Congreso de los Estados Unidos, (registro CIP) se ha utilizado hasta ahora sólo para el material impreso. En el seno del proyecto europeo BIBLINK <<http://www.ukoln.ac.uk/metadata/BIBLINK>> algunas bibliotecas nacionales contemplaron la posibilidad de usar un formato de metadatos simple como el Dublin Core para replicar el proceso CIP para las publicaciones electrónicas, lo que podríamos llamar ECIP. El Dublin Core —como veremos en el capítulo siguiente. *Vid.* 5.2.1.2.— es especialmente adecuado para esta tarea, pues fue concebido originalmente para que el propio autor generase la descripción de un recurso Web. También a lo largo de este trabajo podremos demostrar que la realidad de la asignación de metadatos es mucho más compleja que el ideal de que todo creador de información digital incluya su propia metainformación.

⁴⁹ En inglés, *embed, embedded*. *Cfr.* Glosario. Algunos autores, como Romieux, denominan a este tipo de inclusión de metadatos "catalogación por etiquetas" (*catalogage par balise*). Olivier Roumieux. Balise Meta: indexer ses pages Web. *Archimag*, juillet/août 1998, n° 116, p. 41.

⁵⁰ En este sentido, Weinheimer afirma que *el registro de metadatos debe estar accesible desde el día 1*. James L. Weinheimer. How to Keep the Practice of Librarianship Relevant in the Age of the Internet. *Vine (Theme issue: Metadata. Part 1)*, 1999, n° 116, p 21 y 28.

ejemplo López y Massa⁵¹ opinan que a medida que la cantidad de fuentes de información aumenta, la tendencia a la dispersión puede llegar a crear una situación en la que la búsqueda basada en metadatos tenga tan poco valor como las búsquedas sobre el texto completo (*Cfr.* 7.3).

Sin embargo, el aspecto más controvertido sobre el proceso de creación es, como adelantábamos antes, quién debe ser agente que asigne estos metadatos. Existen varias posturas que nos pueden ayudar a arbitrar sobre este aspecto. Tim Berners-Lee⁵², sin hablar expresamente de quién debe ser el que cree los metadatos, describe de forma sistemática el proceso de creación. Este autor hace hincapié en que los metadatos son, ante todo, datos y con ello quiere decir que la información sobre la información se tiene en cuenta como información en sentido amplio, matizando a través de esta afirmación, que un recurso puede contener datos sobre sí mismo o sobre otro recurso. En el funcionamiento habitual de la World-Wide Web, continúa el inventor de la Web, existen tres caminos a través de los cuales es posible asignar metadatos:

- a) El primero de estos caminos es que los datos sobre un documento estén contenidos dentro de él —por ejemplo en la cabecera de un documento HTML o de un documento XML—. Se obtiene así, como diría Aubrie⁵³, una entidad documento/descripción, en el mismo fichero físico.

⁵¹ Diego R. López y Javier Masa. Dando forma al envase y, con ello, al contenido: Webber [documento HTML]. *Boletín de Red Iris*, octubre 1998, nº 45. Disponible en: <http://www.rediris.es/rediris/boletin/45/enfoque1.html> (consultado el 4 de marzo de 1999).

⁵² Tim Berners-Lee. Metadata Architecture... *Op., cit.*, <http://www.w3.org/DesignIssues/Metadata.html>

⁵³ Claude Aubrie. Les logiciels documentaires. En: *Les nouvelles technologies...* *Op. cit.*, p. 288.

- b) La segunda posibilidad es durante la transferencia de datos, a través del protocolo de transferencia de hipertexto (http), en el cual el servidor puede transmitir también metadatos sobre el objeto que está transfiriendo al cliente.
- c) Finalmente, el tercer camino se produce cuando los metadatos se encuentran registrados en otro documento⁵⁴. Esta tercera opción de Berners Lee, nos hace pensar en la clasificación de los metadatos que apuntábamos en el capítulo anterior (Cfr. 3.5) atendiendo a la fuente de creación que implicaba la posibilidad de asignación de forma externa, es decir por un agente distinto al productor o creador del documento, aunque también pueden ser asignados por la misma entidad creadora, pero almacenados a parte del DLO por cuestiones de la propia arquitectura del sistema de recuperación.

Dempsey y Heery⁵⁵ sí reflejan los posibles autores de los metadatos distinguiendo tres amplias categorías: autores, gestores de depósitos documentales (*repository managers*) y creadores externos (*third party creators*). En primer lugar, los autores son capaces de originar metadatos descriptivos, incentivados además por la existencia, cada vez mayor, de herramientas para la edición y la indización de metadatos utilizando formatos sencillos como el Dublin Core; por otra parte los

⁵⁴ Es oportuno comentar el caso especial de "metadatos" asignados a los DLOs o unidades documentales en la Biblioteca Virtual *Miguel de Cervantes* <<http://www.cervantesvirtual.com>> en la que nos basaremos para proponer un modelo de aplicación práctica en esta tesis doctoral (Capítulo 9). Los documentos textuales, codificados en XML/TEI y publicados en la Web en HTML que forman esta biblioteca tienen, en el código fuente de los documentos visualizados en la Web (HTML) una cabecera de etiquetas <META> prácticamente igual para todos los documentos digitalizados (con la única variación de título-autor en el contenido de la etiqueta <META NAME="title">). Sin embargo existe una especie de "ficha catalográfica" por campos, de la obra original impresa digitalizada, que es también un documento HTML, que a su vez tiene embebido el mismo contenido de metadatos, con la única variación de <META NAME="title" CONTENT="Biblioteca Virtual Miguel de Cervantes">. Estos no son metadatos en sentido estricto, pero sí la información sobre la información (metadatos desde un punto de vista conceptual) de que dispone el lector acerca de los DLOs pertenecientes a ella: el código fuente HTML y la ficha catalográfica. Cfr. 9.3.2.

⁵⁵ Lorcan Dempsey and Rachel Heery. Metadata: A Current View... *Op. cit.*, p. 165.

gestores de depósitos documentales, que tienen cierta responsabilidad sobre los recursos y los datos que describen, podrán también asignar metadatos; finalmente, el establecimiento de metadatos por terceros se producirá cuando los creadores no gestionen ni almacenen los recursos; estos "terceros", pueden ser bien automatizados o humanos, pudiendo incluso estar constituidos por un grupo, como es el caso de las agencias de información que seleccionan y describen de forma unificada, temática y corporativa, recursos creados por otros autores. La creación por un agente externo es sin embargo, indispensable en el caso de la creación de colecciones virtuales distribuidas (*subject gateways*).

Tanto Berners Lee, como Dempsey y Heery reflejan una metodología técnica en términos de posibilidad de asignación; sin embargo, el debate sobre quién debe establecer los metadatos se debe fundar, más que en la posibilidad o factibilidad, en términos de adecuación; es decir, determinar quién es el agente más adecuado, más cualificado o incluso el más barato para la asignación y/o gestión de la metainformación. Existen distintas posturas al respecto, la primera de ellas juega fundamentalmente con el binomio hombre-o-máquina, tradicional en otros ámbitos de la gestión documental y que analizamos a continuación.

Ray Denenberg⁵⁶ propone tanto la opción automatizada como la opción manual para la creación de metadatos como alternativas a los índices globales constituidos por los motores de búsqueda sobre el documento completo:

- En primer lugar, desde el punto de vista de la asignación automatizada de metadatos, señala una arquitectura en la que los propios servidores facilitan información (el segundo camino que señalaba Berners Lee) basada en la teoría de que el servidor es la mejor posición para decidir la indización.

⁵⁶ Ray Denenberg. *Structuring and Indexing the Internet* [documento HTML]. Washington, DC.: Library of Congress, 30 de diciembre de 1996. Disponible en: <http://lcweb.loc.gov/z3950/agency/papers/italy.html> (consultado el 23 de julio de 1998).

Denberg recurre a este modelo porque no sólo mejorará la calidad de la información indizada, como manifiesta la teoría, sino que además reduce la cantidad de información transferida, liberando el tráfico en la Red y el consumo de ancho de banda, problema habitual en la indización en texto completo. Sin embargo el inconveniente de esta propuesta es que requiere mucha cooperación y puede limitar el sistema de indización propietario empleado por muchas compañías.

- Otra alternativa que señala este autor a los índices globales que realizan los motores de búsqueda sobre el texto completo, en este caso, desde el punto de vista de la asignación de metadatos por el hombre, es que los propios creadores de información, los autores o editores de ésta, asignen metadatos para sus documentos (primer camino de Berners Lee).

Catherine Marshall⁵⁷ en una conferencia sobre bibliotecas digitales celebrada en Pittsburgh, se sitúa en una posición clara de alegato en defensa de la asignación de

⁵⁷ Esta autora, parte de un estudio sobre información etnográfica donde se desarrollan metadatos para una colección híbrida formada por recursos visuales, tanto físicos (documentos tradicionales) como digitales (recursos electrónicos, DLOs). Entiende los metadatos desde una perspectiva similar a la que imbuye nuestra investigación: como *la forma de intensificar la utilidad de la información, y concluye que deben ser diseñados para soportar la gestión y el acceso de las distintas, complejas y heterogéneas colecciones que se pueden encontrar en la Red*. Nótese empero, cómo las argumentaciones de Marshall, en pro de una asignación de metadatos por el hombre, mantienen implícita la comparación entre los sistemas basados en metadatos y los sistemas de recuperación sobre el texto completo (tema que abordaremos, desde nuestra propia perspectiva, en el apartado 7.3 de esta tesis), asumiendo que los índices recabados automáticamente por los motores de búsqueda son también metadatos (estamos pues, nuevamente ante un caso de tendencia conceptual del "todo metadatos" que hemos aludido en otras partes de este trabajo y que viene a ejemplificar una vez más la vaguedad o dispersión del término). A pesar de que existe esa comparación implícita entre ambos sistemas, nos ha parecido más interesante traer aquí las ideas desprendidas de la contribución de esta autora en el sentido de que evidencian el interés inicial de que los metadatos sean asignados por el hombre. Catherine C. Marshall. Making Metadata: A Study of Metadata Creation for Mixed Physical-Digital Collection. En: *ACM Conference on Digital Libraries (3. 1998. Pittsburgh)*. Ian Witten, Rob Akscyn, and Frank M. Shipman, eds. New York: ACM, 1998, p. 162-171.

metadatos por el hombre, haciendo prevalecer la idea de Levy⁵⁸ acerca de que los metadatos creados de esta manera *se entienden como una forma de marcar el territorio de nuevos géneros documentales y nuevos medios digitales, y los transforma de tal manera que los convierte en útiles y utilizables*. Los metadatos creados por el hombre —siguiendo a Marshall— son vitales para articular el alcance, el propósito y la función de una colección particular⁵⁹, desde el punto de vista de que pueden realizar una indización orientada al uso de la colección más que orientada al documento en sí mismo, y permiten identificar relaciones entre los propios elementos de una colección (tanto físicos como digitales). Asimismo, los metadatos creados de esta forma, enriquecen las nuevas técnicas de análisis aplicadas a la información multimedia (texto, imagen, audio, vídeo) que describen cada uno de los elementos que forman un DLO, ya que permiten agregar metadatos como palabras clave y/o reflejar la fluidez de los sistemas de clasificación aplicados a un documento estático, que difícilmente se podrían asignar de forma automática.

Para avalar nuestra hipótesis de inexactitud de la asignación automática vamos a utilizar un ejemplo utilizando Klarity <<http://www.klarity.com.au>>, un sistema australiano de la empresa tSA <<http://www.topic.com.au>> para la automatización de la extracción de metadatos a partir de información textual, que justifica la

⁵⁸ D. Levy. Cataloging in the Digital Order. En: *Proceedings of Digital Libraries 95, Austin, TX, June 11-13*, p. 31-37. Citado así en: Catherine C. Marshall. *Ibid.*, p. 162.

⁵⁹ Marshall se refiere a la metainformación asignada en un contexto informativo concreto o biblioteca digital. Anticipa en estas afirmaciones la hipótesis latente en esta tesis doctoral, la utilidad de los metadatos en contextos finitos de información, por ejemplo en colecciones de DLOs temáticos y seleccionados como en el caso de las *subject gateways* (Vid. 8.2.2. y Glosario), donde se prescribe una asignación de metadatos e indización humana ya que la automatización de estos procesos imposibilita la delimitación del contexto y el contenido pragmático de la información de la Red. *Ibid.*

automatización de este proceso en el coste de recursos que implica la creación manual y el riesgo de, aún así, asociar valores incorrectos o poco precisos⁶⁰.

Klarity es un software desarrollado para categorizar los documentos, basándose para ello en los conceptos del texto en lugar de en descriptores. En principio, puede analizar y procesar grandes colecciones de documentos y generar automáticamente metadatos que contribuyan tanto a la clasificación como a la localización de la información, pudiendo adaptar el formato de salida a varios modelos de metadatos (etiquetas <META> de HTML, RDF y texto ASCII). Para analizar las capacidades de Klarity como sistema completamente automatizado de creación elegimos un documento, predominantemente textual, justamente relacionado con el tema de la asignación automática de metadatos y que comentaremos a continuación: *Automatic RDF Metadata Generation for Resource Discovery*⁶¹, cuyo URL es: http://www.scit.wlv.ac.uk/~ex1253/rdf_paper. Este es el resultado inmediato que da Klarity sobre los metadatos de dicho documento (*tilte, keyword, description*):

⁶⁰ Podríamos profundizar más aquí en el eterno debate entre las ventajas e inconvenientes de la indización automatizada frente a la indización humana. Esta cuestión la hemos tratado ya en: José Antonio Moreiro González, Eva M^a Méndez Rodríguez. Lenguaje natural e indización automatizada. *Ciencias de la Información*, septiembre 1999, vol. 30, n^o 3, p. 11-15. Sobre este aspecto, Klarity hace un defensa en pro de la automatización de la creación de metadatos, sin embargo, y como demuestra el ejemplo a continuación, veremos que no se puede confiar completamente a un software la asignación y categorización del contenido de un DLO.

⁶¹ Hemos elegido una página Web en inglés, en lugar de una en español para no limitar la capacidad de Klarity en el procesamiento del lenguaje natural, que entendemos que será mayor en inglés.

El resultado http://www.klarity.com.au/?session=browser&action=klarity&url=http://www.scit.wlv.ac.uk/~ex1253/rdf_paper aparece al ejecutar el programa sobre una página Web simplemente pulsando el enlace que previamente hemos colocado en la barra personal de nuestro navegador, según indica Klarity en: http://www.klarity.com.au/?action=showpage&session=135996064&file=klarity_add

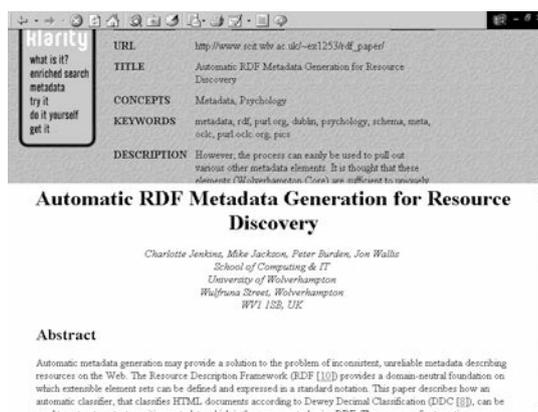


Fig. 3. Creación automática de metadatos con Klarity

Y estos mismos metadatos, utilizando la sintaxis RDF, según la opción que presenta el programa, son:

```
<?xml version='1.0'?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc = "http://purl.org/dc/elements/1.0/"
xmlns:dcq = "http://purl.org/dc/qualifiers/1.0/">
<rdf:Description rdf:about =
"http://www.scit.wlv.ac.uk/~ex1253/rdf_paper/">
<dc:description>However, the process can easily be used to pull out
various other metadata elements. It is thought that these elements
(Wolverhampton Core) are sufficient to uniquely identify the
document, state where it can be found, provide a good indication of
the subject matter and of how current both the actual information
and its metadata are. The implications of this comparison are
discussed again in the later section (3.2) on RDF schema
definition.</dc:description>
<dc:subject>
<rdf:Description>
<rdf:value>Metadata (1.67655), Psychology (1.06122)</rdf:value>
<dcq:subjectType>classification</dcq:subjectType>
<dcq:subjectScheme>Klarity</dcq:subjectScheme>
</rdf:Description>
</dc:subject>
<dc:subject>metadata, rdf, purl.org, dublin, psychology, schema,
meta, oclc, purl.oclc.org, pics</dc:subject>
<dc:title>Automatic RDF Metadata Generation for Resource
Discovery</dc:title>
</rdf:Description>
</rdf:RDF>
```

En cuanto a la automatización del etiquetado, siguiendo la sintaxis RDF y el modelo DC, no podemos hacer ninguna objeción puesto que son tareas perfectamente automatizables. Sí queremos llamar la atención sobre el análisis de contenido que realiza, esto es, las materias extraídas (conceptos) y el resumen o descripción del documento. Es importante mencionar también que no extrae ninguna información

sobre la autoría del documento, dato en cualquier caso indispensable en una estructura de metadatos. En cuanto al resumen está formado por tres fragmentos no continuos extraídos literalmente del texto que no condensan el contenido real del documento. Los descriptores (o conceptos según la propia terminología de Klarity), si bien son términos que aparecen en el documento, salvo en el caso de *metadata* y *rdf*, no caracterizan el contenido del mismo. Por último, los valores de clasificación — Metadata (1.67655), Psychology (1.06122)— indican una notación numérica que responde al esquema de clasificación propietario de Klarity y que en el caso de *psychology*, a pesar de que aparezca trece veces a lo largo de los ejemplos consignados en el documento, no responde en absoluto al contenido temático de éste. Estos resultados anómalos o imprecisos pueden deberse a la propia extracción de metadatos basada en conceptos⁶² que utiliza esta herramienta.

Contrariamente a lo que demuestra el funcionamiento de una herramienta de asignación automática como Klarity, algunos investigadores de la Universidad británica de Wolverhampton defienden la generación automática de metadatos RDF para la recuperación de información⁶³. La visión de estos autores es muy realista al

⁶² Vid. tSA Consulting Group. Concept vs Keywords [documento HTML]. En: *Automatic Generation of Metadata Based on Concepts Within the Document* [meta-data whitepaper]. Klarity.com, 4 de Julio de 2000. Disponible en: http://www.klarity.com.au/whitepaper/meta-data_white_paper-06.htm#P70_7248 (consultado el 21 de marzo de 2001). A pesar de que, ante nuestras preguntas acerca del funcionamiento interno de Klarity, la empresa se reserva algunas cuestiones en el sentido de que pertenecen a la propiedad intelectual del programa, podemos decir que, en líneas generales, Klarity pasa por diversas fases antes de presentar los metadatos que automáticamente asigna a una página: en primer lugar determina una taxonomía o vocabulario controlado para una materia; después identifica los documentos ejemplo analizando densidades y otros parámetros para desarrollar el fichero de concepto que relaciona los conceptos con sus significados; la fase siguiente es el análisis del nuevo contenido, puntuando o calificando cada término de la taxonomía, según esto se pueden ordenar grandes cantidades de documentos basándose en cada elemento de la taxonomía; finalmente, una vez que se determinan los metadatos, el programa utiliza una plantilla para visualizarlos en el formato indicado (en este caso RDF). Graeme Cox. Regarding your Klarity Enquiry [correo electrónico personal] 28 de marzo de 2001; 02:33. Disponible en: emendez@bib.uc3m.es

⁶³ Charlotte Jenkins, et al. Automatic RDF Metadata Generation for Resource Discovery [documento HTML]. En: *International World Wide Web Conference (8. 1999. Toronto)*, rev. 5 de diciembre de [cont.]

tener en cuenta que actualmente la mayoría de los creadores no incluyen metadatos en los recursos que publican. Si bien es difícil automatizar la descripción de recursos sería imposible describir cada uno de ellos manualmente. Por todo ello proponen un generador automático de metadatos para habilitar descripciones RDF asociadas a páginas HTML independientemente de cuándo y cómo (con qué herramientas) hayan sido creadas éstas.

A pesar de la justificación de estos autores, y de la lógica de la automatización de la asignación de metadatos dado el tamaño de la WWW, hoy en día todas las experiencias de creación automática de la información sobre la información pertenecen al ámbito de la investigación y distan mucho de ser una realidad práctica generalizada en todos los sistemas de información. Así pues, parece estar claro que la creación de metadatos se debe confiar al hombre en lugar de a un sistema exclusivamente automatizado, máxime si se trata de formar colecciones específicas consistentes en un universo de información heterogéneo. Sin embargo, creemos que en los metadatos, como en otras funciones tradicionalmente asignadas al bibliotecario como mediador de información, el éxito como sistema para mejorar la recuperación de información estará basado en una interacción entre el hombre y los agentes de software. Esta idea la defiende Zick⁶⁴, no estrictamente en el ámbito de los metadatos pero sí extrapolable a ellos, al señalar las diferencias y las semejanzas entre los bibliotecarios y los agentes de software. Esta autora señala que tanto el trabajo de los bibliotecarios como de los agentes automáticos se desarrolla en un mundo incierto y tanto unos como otros tendrán que enfrentar los mismo objetivos, retos y problemas.

2000. Disponible en: <http://www8.org/w8-papers/2c-search-discover/automatic/automatic.html> (consultado el 18 de marzo de 2001). El software que han desarrollado para la creación automática de metadatos en formato RDF, así como para la clasificación según la DDC se denomina *Automatic RDF Metadata Generator*, y se puede utilizar desde: <http://scitsd.wlv.ac.uk:8080/metadata.html>; o desde <http://www.scit.wlv.ac.uk/~ex1253/metadata.html>

⁶⁴ Laura Zick. The Work of Information Mediators: A Comparison of Librarians and Intelligent Software Agents [documento HTML]. *First Monday*, May 2000, vol. 5, nº 5. Disponible en: http://www.firstmonday.dk/issues/issue5_5/zick/index.html (consultado el 15 de mayo de 2001).

Asimismo Lynch⁶⁵ predice que todo esquema de metainformación adoptado por la comunidad científica necesita dos componentes: *generación automática de los metadatos y anotaciones generadas por el hombre*.

En esta línea, Desai⁶⁶ propone una alternativa mixta a la creación de metadatos: la asignación de encabezamientos semánticos por el hombre pero mediante un sistema experto, donde el sistema de indización responde a una arquitectura distribuida y accesible, en este caso, tanto por los proveedores como por los usuarios de Internet. Desai⁶⁷ parte de que en un sistema de información distribuida como Internet, es natural que los proveedores de información preparen e introduzcan la información bibliográfica (metadatos) sobre cada recurso usando un esquema de indización estandarizado. Hasta aquí, la teoría de este autor coincide con todo lo que hemos argumentado hasta el momento, esto es, la creación de metadatos por el agente productor de la información (autor/editor) siguiendo un modelo concreto, sencillo y más o menos normalizado como pudiera ser el Dublin Core. Sin embargo, lo novedoso de la aportación de este autor es que estos sistemas deben incorporar al

⁶⁵ Las predicciones de Lynch van en consonancia con el propio funcionamiento que hemos visto en Klarity: es necesaria la automatización de la producción de metadatos, sin embargo el contenido de las anotaciones de las etiquetas debe de realizarse por el hombre. Clifford Lynch. *Op. cit.* <http://sciam.com/0397/issue/0397lynch.html>

⁶⁶ Los trabajos de este investigador canadiense en esta línea se pueden consultar en: Bipin C. Desai. Supporting Discovery in Virtual Libraries. *Journal of the American Society for Information Science*, 1997, vol. 48, nº 3, p. 194 y ss.; Bipin C. Desai. *The Semantic Header and Indexing and Searching on the Internet* [documento HTML]. Montreal: University of Concordia, Department of Computer Science rev. 11 de agosto de 1995. Disponible en: <http://www.cs.concordia.ca/~faculty/bcdesai/cindi-system-1.0.html> (consultado el 18 de junio de 1998); Bipin C. Desai. *The Semantic Header and Indexing and Searching on the Internet* [documento HTML]. Montreal: University of Concordia, Department of Computer Science, febrero de 1995, rev. 30 de septiembre de 1997 en: <http://www.cs.concordia.ca/~faculty/bcdesai/cindi-system-1.1.html> (consultado el 9 de agosto de 2001). Lamentablemente los trabajos de Desai en torno al SH no han evolucionado demasiado desde su fecha de publicación, aunque sí es destacable la originalidad y adecuación de sus ideas "para la época" (1995-97).

⁶⁷ Bipin C. Desai. Supporting Discovery... *Ibid.*, p. 192.

menos, la experiencia y el conocimiento básicos de los catalogadores y los bibliotecarios de referencia a través de un sistema de ayuda (sistema experto) que guíe al usuario en todos los pasos. El sistema de entrada de índices y el sistema de búsqueda (clientes) deben ser locales tanto para el proveedor como para el usuario de la información, mientras que el sistema de base de datos bibliográfica (cliente) debe distribuirse y reproducirse en un número de nodos regionales para mejorar el acceso y la respuesta.

El modelo de Desai, se origina como revulsivo al problema de la pobreza selectiva de los actuales índices de búsqueda en Internet, y crea un formato llamado encabezamiento semántico (*Semantic Header=SH*⁶⁸). El SH de Desai, no es más que un modelo más de metadatos que podríamos analizar en el apartado dedicado a ellos (capítulo 5), pero lo tratamos ahora ya que, si bien como formato de metadatos no es más que otra alternativa de estructuración de la información para su descripción, en cuanto al proceso de creación supone una particularidad interesante en la inclusión de un subsistema experto fundamentado en bases de conocimiento. Este subsistema guía al usuario en la elección de términos estandarizados a través de una interfaz gráfica fácil de usar. Con ello podríamos decir que la asignación de metadatos según el modelo del SH de este autor, responde a un proceso de creación mixto: usuario proveedor/autor de la información + sistema automático experto (imbuido del conocimiento y la experiencia del catalogador profesional). Es decir, el SH implica una estructura que permite una extracción de metadatos semiautomática, asistida por el hombre, constituyendo además un sistema de indización distribuido.

⁶⁸ El encabezamiento semántico de Desai es un intento más de normalización de la información bibliográfica de la Red para permitir, fundamentalmente el intercambio y la recuperación de información. Su sintaxis se basa en el lenguaje de marcas SGML (etiqueta <SEMHDR>), e incluye aquellos elementos más comúnmente usados en la recuperación de información: título, título alternativo, autor y otras menciones de responsabilidad, materia y submateria (palabras clave e identificadores), fecha, versión y clasificación. Bipin C. Desai. Supporting Discovery... *Op. cit.*, p. 195-199.

No obstante, sea con ayuda de aplicaciones informáticas o sin ellas, consabida la creación por el hombre, debemos precisar ahora otro aspecto fundamental: qué agente humano concreto creará tal información: ¿el propio autor del documento?, ¿el editor?, ¿personal cualificado/catalogador? La opinión difundida en el *Warwick Framework*⁶⁹ (segundo taller del Dublin Core) sobre esta cuestión parte del coste de creación de los metadatos, especialmente cuando se generan automáticamente, lo que ha llevado a investigaciones sobre la posibilidad, ventajas y desventajas de que el autor o editor asignen los metadatos cuando éstos se utilizan dentro del propio documento (en la cabecera de un documento HTML, SGML o XML). Conforme a esa apreciación, los esquemas para que puedan crearse por no-especialistas tienen que ser muy sencillos, cortos, y fáciles de entender y usar. El hecho de que los documentos de la Red sean cada vez más numerosos, y su indización a través de sistemas automatizado sea costosa, obliga a que sean los autores/editores los que creen los metadatos para la mayoría de los DLOs y la existencia de aplicaciones sencillas ayudará a la viabilidad de esta opción.

A pesar de que esta alternativa, que sea el propio autor del DLO el que asigne la metainformación como parte estructural a la creación de documentos para la Web, parece ser la más aceptada y la que de forma tácita venimos reafirmando a lo largo de todo este apartado, existen opiniones contrarias. Dicho de otra forma, es preciso también considerar algunos riesgos que conlleva la responsabilidad de creación de los metadatos por parte de los autores. Denenberg⁷⁰ considera que la creación de *metadatos de autor* no es una opción práctica por múltiples razones: resulta muy divergente a los esquemas de indización propietarios utilizados por las empresas que

⁶⁹ Juha Hakala, Ole Husby and Traugott Koch. *Warwick Framework and Dublin Core Set Provide a Comprehensive Infrastructure for Network Resources: Report from the Metadata Workshop II, Warwick, UK, April 1-3, 1996* [documento HTML]. Lunds University Electronic Library, 6 de abril de 1996, rev. 19 de junio de 1996. Disponible en: <http://www.ub2.lu.se/tk/warwick.html> (consultado el 28 de julio de 1998).

⁷⁰ Ray Denenberg. *Op. cit.*, <http://lcweb.loc.gov/z3950/agency/papers/italy.html>

desarrollan Sistemas de Recuperación de Información de Internet (SRII); requiere mucho trabajo por parte de los autores que no están, en muchos casos, dispuestos a realizar, y en el caso de que lo estén, podrían hacerlo mal y, asignar metadatos erróneos es peor que no incluir ninguno; otro de los riesgos de los metadatos de autor es la tentación de añadir términos con el simple propósito de incrementar el orden de relevancia potencial del documento.

Aunque advertimos los mismos riesgos que apunta Denenberg, es el momento de diferenciar los dos contextos de recuperación de información a los que nos referimos al principio:

- Por un lado, en el contexto general de búsqueda en la WWW a través de motores y herramientas automáticas de indización, clasificación y búsqueda. En este ámbito está claro que es el autor/editor el que mejor puede crear los metadatos ya que son los que más cerca se encuentran del momento "cero" del documento, y resulta menos costoso que crear un sistema automatizado o pagar personal cualificado; si bien es cierto que, para que el autor/editor cree su propia metainformación tiene que contar con un formato fácil y estar incentivado, como señalaban Dempsey y Heery, por la existencia de las herramientas necesarias para seguir fácilmente un modelo concreto de metadatos.
- Por otro lado, Michel Day⁷¹ se refiere al contexto de los sistemas de información de calidad (*subject gateways*), presentando un elenco de opciones para los creadores de metadatos a parte de que también señala la opción de crearlos automáticamente. Según este autor, en este entorno finito de información podrían proporcionar metadatos de los recursos Web: los proveedores de información, los

⁷¹ DESIRE. *Information Gateways Handbook (Print Version)* [documento HTML]. DESIRE, rev. 26 de abril de 2000. Disponible en: <http://www.desire.org/handbook/print4.html>, [p. 38 en la versión impresa, o en el URL: <http://www.desire.org/handbook/2-3.html>] (consultado el 1 de junio de 2000).

usuarios del servicio de información, catalogadores especializados, editores, un grupo seleccionado de personal u otro servicio de información semejante que colabore. En la misma línea, y fijándonos en sistemas de metadatos concretos, que también responden a nuestra teoría de sistemas de información con propósitos específicos dentro de un contexto finito de recursos Web, las distintas guías de uso de esquemas determinados⁷² también hacen referencia a posibles creadores de metadatos en estos términos: la persona que crea el recurso, el responsable de la publicación, el personal bibliotecario empleado para catalogar los documentos, o intermediarios independientes, tales como una agencia particular del gobierno, una biblioteca o un proveedor de servicios comerciales.

Al menos como una opción, en general todos los autores coinciden en que puede o debe ser el autor del documento el creador de metainformación. En el contexto más definido de las bibliotecas digitales y de las *subject gateways* y *clearinghouses*⁷³, los responsables tenderán a ser especialistas, desde bibliotecarios a agencias especiales, llegando a ser más exigentes en el caso de los servicios temáticos de información geoespacial⁷⁴, donde se aconseja que sean bien *los administradores o gestores del sistema que manejan conocimientos científicos, o bien los científicos que manejan con soltura conceptos específicos de informática*. En estos sistemas de información especializados, donde prevalece una selección y

⁷² Por ejemplo la guía de BEP: The Creation of Metadata [documento HTML]. En: *Metadata Specification for BEP, Business Entry Point Management Branch of Department of Employment, Workplace Relations and Small Business*, 1999. Disponible en: <http://about.business.gov.au/bep/agencies/provinfo/metadata/meta02.htm> (consultado el 28 de febrero de 2001).

⁷³ Vid. Glosario.

⁷⁴ *Frequently-Asked Questions on FGDC Metadata* [documento HTML]. Maintained by Peter Schweitzer. USGS, rev. 3 de julio de 2000. Disponible en: <http://geology.usgs.gov/tools/metadata/tools/doc/faq.html#1.3> (consultado el 2 de junio de 2001); y *Clearinghouse: Preguntas más frecuentes sobre metadatos* [documento HTML]. Montevideo: Clearinghouse Nacional de Datos Geográficos, rev. 16 de diciembre de 1999. Disponible en: <http://www.clearinghouse.com.uy/metadatos/pregunta.html> (consultado el 23 de julio de 2000).

organización de la información, se compara la creación de metadatos con la catalogación bibliotecaria, salvo en que el creador, en este caso necesita además saber más sobre la información específica que hay detrás de los datos de tal forma que puedan describirlos apropiadamente. Con todo, el personal especializado o los "catalogadores" contribuirán con un porcentaje más pequeño de descripciones de recursos y su trabajo probablemente se concentrará en los documentos más duraderos o de mayor calidad, o para describir recursos electrónicos de sistemas de recuperación concretos y específicos basados en metadatos, que utilicen modelos más complejos o de propósito individual.

Para avalar esta última opinión, Linda Hill⁷⁵ argumenta que en el caso de la *Alexandria Digital Library* (ADL⁷⁶) es el personal bibliotecario profesional quien crea los metadatos para las colecciones de ítems de la biblioteca de mapas e imágenes propias, usando un software basado en Access y desarrollado localmente; además, en esta biblioteca digital, para la recuperación pretenden integrar recursos basados en otros modelos de metadatos, por lo que han mapeado la metainformación creada de forma externa por otras instituciones (en formato MARC o en formatos de metadatos contruidos *ad hoc* por ellas) al esquema de metadatos de la ADL o directamente a sus *search buckets*⁷⁷ que proporcionan parámetros comunes para la búsqueda entre

⁷⁵ Linda Hill. Re:NKOS: Who Assign Metadata? [correo electrónico en lista de distribución] NKOS, 29 de julio de 1998; 17:02 [no existe archivo de la lista consultable de esa fecha]. Disponible en: emendez@bib.uc3m.es

⁷⁶ Recordemos que la ADL <<http://www.alexandria.ucsb.edu>> es uno de los seis proyectos de bibliotecas digitales de la *National Science Foundation* (NSF) (DLI1, 1994-1998), desarrollado por la Universidad de California en Santa Bárbara para crear una biblioteca digital especializada (Vid. Capítulo 1, nota 10), en definitiva, un contexto finito de información, en este caso, temática. Desde 1999 este proyecto se ha convertido en *Alexandria Digital Earth Prototype* <<http://www.alexandria.ucsb.edu/adept>> (DLI2, 1999-. Cfr. 8.2.1.), que va más allá de la metáfora de la biblioteca, a la metáfora de "la tierra digital"; tratando de dar acceso a recursos de información heterogéneos y distribuidos a través de lo que denominan *Isclapes* (*Information landscapes*, paisajes de información).

⁷⁷ Vid. Glosario.

colecciones heterogéneas. Hill matiza además que en la ADL concretamente prefieren convertir los metadatos existentes más que crearlos *ex-novo*. De esto último se deduce también que, ante la posibilidad de adaptar descripciones hechas previamente, y dado el número exponencial de documentos (DLOs) contenidos en la Red, se valora el tiempo del especialista, a la vez que esta actitud puede beneficiar comportamientos cooperativos en el entorno de Internet para la descripción oportuna de recursos electrónicos en un área concreta. Esta valoración de la interoperabilidad a tenor de la globalización de la información, encaja por otra parte como veremos, en el contexto del *Warwick Framework*, que encomia los formatos de metadatos legibles por ordenador y perfectamente operativos o mapeables en otros sistemas de información ya que, según el pensamiento emanado de este encuentro de la DCMI, la mejora de la calidad y precisión de las herramientas de búsqueda y recuperación de información automáticas en Internet depende del acopio de formatos más ricos⁷⁸.

Weinheimer⁷⁹ concilia ambas posturas (la de la asignación de metadatos por los creadores de la información y la de los catalogadores) haciendo un paralelismo con las tareas básicas de la catalogación: descripción y asignación de puntos de acceso. Es responsabilidad de los creadores de las páginas Web, la información descriptiva, es decir, en atributos del DC, título, descripción, editor, fecha, tipo de recurso, identificador, idioma, cobertura y derechos; y revierte en los "catalogadores" la responsabilidad de los puntos de acceso, esto es, autor, materia, relación y fuente. A pesar de lo conciliador de esta postura sobre la experiencia que debe tener el agente que asigna los metadatos y de la coherencia de la propuesta, seguimos pensando que la determinación del quién debe crear la metainformación, depende del tipo de recurso Web que sea y del tipo de servicio responsable de su recuperación: el

⁷⁸ Juha Hakala, Ole Husby and Traugott Koch. *Op. cit.*, <http://www.ub2.lu.se/tk/warwick.html>

⁷⁹ James L. Weinheimer. *Op. cit.*, p. 25.

autor, en el caso de la búsqueda genérica, y personal especializado en el caso de las bibliotecas digitales.

En último lugar, nos resta por tratar otro tema estrechamente vinculado a la asignación de los metadatos, e incluso a los responsables de su creación, nos referimos a su arquitectura y almacenamiento. Aunque en cierto sentido hemos mencionado tácitamente, la forma más adecuada de almacenamiento, al decir que los metadatos deben estar embebidos en el mismo documento, y que al clasificarlos consignábamos dos tipos fundamentales según este criterio (metadatos embebidos y *stand-alone metadata*⁸⁰), debemos contemplar también otras opciones tanto desde el punto de vista del formato y sintaxis, como del almacenamiento en sí mismo. Así, los metadatos se pueden guardar, principalmente en⁸¹:

- 1) El propio objeto/documento (DLO) descrito, normalmente como cabecera. Esta forma de almacenamiento tiene la ventaja de que la información está contenida en el documento y puede transportarse a través del sistema. Las opciones de sintaxis para este modelo de almacenamiento son principalmente dos:
 - Metaetiquetas HTML asociadas directamente con el recurso y por tanto accesibles a los motores de búsqueda generales. Esta es una forma de codificación adecuada para sitios Web pequeños e independientes, sin

⁸⁰ Basándonos en el criterio de la DCMI (Dublin Core Metadata Initiative). *Cfr.* Capítulo 3, nota 121. *Vid.* 5.2.1.2.

⁸¹ Para hablar de las formas principales de almacenamiento de metadatos nos hemos basado sobre todo, en: Chris Taylor *Op. cit.*, <http://www.library.uq.edu.au/iad/ctmeta4.html>. BEP: The creation of metadata. *Op. cit.* <http://about.business.gov.au/bep/agencies/provinfo/metadata/meta03.htm>; Klarity: Where do I store Metadata?; Klarity: What Format do I Choose for Storage? [documentos HTML]. En: *Automatic generation of metadata based on concepts within the document [meta-data whitepaper]* [documento HTML]. Klarity.com, 4 de Julio de 2000. Disponible en: http://www.klarity.com.au/whitepaper/meta-data_white_paper-09.htm#P111_10124 (consultado el 26 de marzo de 2001); y *DCMI Frequently Asked Questions (FAQ)* [documento HTML]. DCMI, rev. 9 de febrero de 1999. Disponible en: <http://dublincore.org/resources/faq> (consultado el 17 de marzo de 2001).

embargo, para sitios grandes y complejos, que mantienen relación con otros, tiene el inconveniente de que sólo soporta una lista sencilla de metaelementos, sin posibilidad de agrupación ni de otra estructura, además, en la codificación en HTML resulta difícil la actualización y mantenimiento de las etiquetas <META>⁸².

- Etiquetas RDF/XML almacenadas en este caso, bien en el propio recurso, bien de forma separada. La codificación XML es una forma mucho más flexible que las metaetiquetas HTML, sobre todo cuando los metadatos se almacenan aparte del documento, ya que permite una mayor estructuración y aporta mecanismos como los *namespaces*⁸³ que añaden valor a esta metainformación al permitir combinar varios esquemas. El mayor problema que tiene esta sintaxis de almacenamiento es que todavía las herramientas de software que procesan XML/RDF para la búsqueda, son escasas y en fase de experimentación.
- b) Como hemos anticipado, el almacenamiento de los metadatos puede ser también en un fichero aparte al que se accede de forma externa y que, por lo general, apuntan o se dirigen al documento u objeto de información. Son los metadatos que Murphy⁸⁴ denomina "separables", y los define como *aquellos que pueden almacenarse independientemente del documento* y por tanto persisten aunque el documento ya no esté accesible. Dicha autora distingue entre los metadatos

⁸² Existe una especificación del IETF, en forma de *Request for Comments* (RFC 2731) <<http://www.ietf.org/rfc/rfc2731.txt>> sobre la codificación del modelo DC en el elemento <META> de la cabecera HTML. También hablaremos del almacenamiento de metadatos en HTML, al hablar de la metainformación de propósito general (Vid. 5.2.1.1).

⁸³ Vid. 5.2.3 y Glosario.

⁸⁴ Lisa D. Murphy. Addressing the Metadata Gap: *ad hoc* Digital Documents in Organizations. En: *Text Databases and Document Management: Theory and Practice*. Amita Goyal Chin, ed. Hershey, etc.: Idea Group Publishing, 2001, p. 55.

físicamente separables y los separables lógicamente; estos últimos son elementos de información que no se pueden determinar a través del acceso directo al contenido del DLO, tales como la fecha de nacimiento y muerte del autor o la existencia de distintas versiones del documento. Ejemplos de metadatos, en sentido amplio, separables son, por ejemplo, los índices generados automáticamente por aplicaciones como Word2000 o las descripciones presentadas en los portales como Yahoo⁸⁵.

Esta forma es fácil de aplicar y de indizar, pero tiene el problema de que todavía son escasas las herramientas de gestión integradas que permitan diseñar sistemas de recuperación de información basados en metadatos externos. Esta manera de almacenamiento es especialmente aconsejable para describir imágenes⁸⁶ o DLOs multimedia. La manera más fácil de crear metadatos externos y asociarlos a una página Web es a través del propio elemento LINK del HTML⁸⁷, de forma semejante a como se vinculan las hojas de estilo (CSS) externas.

⁸⁵ *Ibid.* Estos metadatos son los que crean herramientas como Metacontent Builder. *Cfr. infr.* 4.3.2.2.

⁸⁶ No obstante, existe una aplicación, RDFPic <<http://jigsaw.w3.org/rdfpic>>, desarrollada por el W3C, para embeber una descripción RDF de una foto, en la propia imagen, tal y como se expresa en una Nota del Consorcio Web sobre este tema: *Yves Lafon, Bert Bos. Describing and Retrieving Photos Using RDF and HTTP: W3C Note* [documento HTML]. W3C, 28 de septiembre de 2000. Disponible en: <http://www.w3.org/TR/photo-rdf> (consultado el 6 de octubre de 2000).

⁸⁷ De esta forma están, por ejemplo, los metadatos asociados a las páginas Web de la DCMI que, en sentido estricto, no es más que una página normal de Internet (codificada en XHTML) cuya temática son los metadatos. Por ejemplo, si visualizamos el código fuente de la página: <http://dublincore.org>, encontramos <LINK REL="meta" HREF="index.shtml.rdf"/>, que es la información que relaciona el recurso <<http://dublincore.org>>, con sus metadatos <<http://dublincore.org/index.shtml.rdf>>:

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc = "http://purl.org/dc/elements/1.1/">
<rdf:Description rdf:about="http://dublincore.org/">
<dc:title>Dublin Core Metadata Initiative (DCMI) Home Page</dc:title>
<dc:description> The Dublin Core Metadata Initiative is an open
forum engaged in the development of interoperable online metadata
standards that support a broad range of purposes and business
models. DCMI's activities include consensus-driven working groups,
```

[cont.]

- c) La tercera forma de almacenamiento es en una base de datos que apunta al DLO. El hecho de almacenar los metadatos de esta forma permite sistemas de gestión más sofisticados. Es la forma más compleja y cara de implementar la tecnología de metadatos para la recuperación de información, pero también la más flexible y rentable ya que tiene dos ventajas principales: múltiples conjuntos de metadatos pueden referirse al mismo recurso de tal forma que atiendan a las necesidades de distintas comunidades de usuarios, y por otra parte, el administrador tiene la posibilidad de parametrizar y adaptar las funciones de una base de datos. Esta forma de almacenamiento es la que sigue por ejemplo *MetaStar Repository*, uno de los módulos que incluye el "software de gestión integrada de bibliotecas digitales" desarrollado por Blue Angel⁸⁸.

La metainformación en una base de datos se almacena separada de los DLOs (bien por campos, bien guardando todo el conjunto de metadatos asociados a un documento en un registro), pudiendo estar en el mismo sitio que el recurso o en otro; además, estas bases de datos no son accesibles por los motores de búsqueda públicos y requieren la definición de un API (*Application Program Interfarce*) para hacerse utilizables, como por ejemplo, en nuestro ámbito más bibliotecario, Z39.50. De cualquier forma, almacenar metadatos en una BD es mejor en términos de actualización de los datos y de mantenimiento, conversión a otro formato o migración a estándares futuros.

Los sistemas de gestión de bases de datos consisten básicamente en varios campos interrelacionados de alguna manera que son los más adecuados, en principio, para almacenar y recuperar datos, lo son también para almacenar

```
global workshops, conferences, standards liaison, and educational
efforts to promote widespread acceptance of metadata standards and
practices.</dc:description> <dc:date>2001-01-16</dc:date>
<dc:format>text/html</dc:format> <dc:language>en</dc:language>
<dc:contributor>Dublin Core Metadata Initiative</dc:contributor>
</rdf:Description> </rdf:RDF>
```

⁸⁸ En el apartado 4.3.2.3 nos referiremos a este software. Aunque destacaremos su forma de creación de metadatos, obviamente también tiene un sofisticado sistema de almacenamiento para los mismos.

metadatos. El modelo y la sintaxis de RDF, como veremos en el capítulo siguiente, son muy similares a esta sencilla definición de SGBD, por ello no es sorprendente que se traten de aparejar las funcionalidades de las bases de datos a los metadatos codificados en RDF. A pesar de que estas afirmaciones son lógicas desde el punto de vista de la teoría, distan mucho de ser una realidad. Actualmente, a raíz de la lista de discusión RDF-interest⁸⁹ del Consorcio Web, se están debatiendo distintas formas de almacenar RDF en bases de datos relacionales⁹⁰, sin embargo, estas aplicaciones están aún en fase de desarrollo.

- d) Otra manera de almacenar es una variedad de la anterior (bases de datos): un sistema de gestión de registros donde se crea la metainformación directamente en una base de datos y se enlaza con el recurso. Esta modalidad es la más apropiada en el caso de que se quiera realizar una biblioteca digital directamente con recursos seleccionados de la Web, donde la biblioteca en cuestión no está implicada en el proceso de producción del documento digital. La creación de metadatos por un agente externo y su almacenamiento en una base de datos, es

⁸⁹ Semantic Web Activity: RDF Interest Group <<http://www.w3.org/RDF/Interest>>. La suscripción a esta lista de discusión desde Agosto de 1998, nos ha permitido hacer un seguimiento tanto del desarrollo de RDF, como de las iniciativas que surgen para su almacenamiento.

⁹⁰ Sergey Melnik ha recogido todas estas aportaciones en: *Storing RDF in a Relational Database* [documento HTML]. Stanford: Stanford University, Computer Science Department, 11 de mayo de 2000. Disponible en: <http://www-db.stanford.edu/~melnik/rdf/db.html> (consultado el 1 de octubre de 2000). Posteriormente Barstow tratando de consolidar la información sobre los sistemas utilizados para almacenar metadatos RDF ha creado un recurso donde se recoge la descripción de varios de ellos. Art Barstow. Survey: RDF Data Stores [correo electrónico en lista de distribución]. *www-rdf-interest*, 19 de abril de 2001; 15:04. Disponible en: <http://lists.w3.org/Archives/Public/www-rdf-interest/2001Apr/0327.html> (consultado el 22 de abril de 2001). Más reciente es la contribución, en este mismo sentido, de Nelson Henrique Zanete. *A Contribution to Brian McBride's Approach* [documento HTML]. Porto Alegre: FACCAR, Universidade Federal de Rio Grande do Sul, 6 de julio de 2001. Disponible en: <http://www.faccar.com.br/zanete/rdfdb.htm> (consultado el 13 de julio de 2001). Hjelm también recoge y comenta algunas de las principales tentativas de almacenamiento de RDF en bases de datos. Johan Hjelm. *Creating the Semantic Web with RDF: Professional Developer's Guide*. New York, etc.: Wiley Computer Publishing, John Wiley & Sons, 2001, p. 261-262. Por último también es interesante destacar la iniciativa XML:DB <<http://www.xmldb.org>> que pretende desarrollar la tecnología para la gestión de datos y metadatos en bases de datos constituidas en XML.

una fórmula creciente sobre todo para el desarrollo de *subject gateways*⁹¹ (por ejemplo la *Australian Virtual Engineering Library* <<http://avel.edu.au>>). En este sentido también hay que tener en cuenta las soluciones de los grandes SIGB (Sistemas Integrados de Gestión de Bibliotecas) en el tratamiento y gestión de metadatos, tendentes sobre todo a asegurar la integración de registros de metadatos y de registros convencionales⁹².

A pesar de todas estas posibilidades, la elección de la forma de almacenamiento tiene consecuencias en la creación y funcionamiento de un sistema de información basado en metadatos. Verbigracia, la utilización de metadatos integrados requiere herramientas que puedan distinguir que están embebidos en el propio código del documento y que sean capaces de manipularlos tanto para la búsqueda como para la visualización. Para los metadatos separados del documento, sin embargo, los problemas están relacionados con la posibilidad de asociarlos con el objeto de información al que se refieren, incluyendo en esta forma de gestión, la

⁹¹ Vid. 8.2.2., Glosario.

⁹² La tendencia de estos SIGB es facilitar la creación de bibliotecas híbridas (Cfr. 8.3.2.), pero todavía son pocos los que permiten la creación y almacenamiento de metadatos, más allá de la inclusión del campo 856 del MARC. Sin embargo están empezando a aparecer herramientas específicas para la gestión de distintos formatos de metainformación en entornos integrados de gestión de bibliotecas. Este es el caso por ejemplo de Hyperion de Sirsi y MetaLib de Ex Libris (Cfr. 4.3.2.3.):

- Sirsi, la empresa que ha desarrollado y comercializa Unicorn, plantea la gestión de colecciones digitales a través de un módulo denominado Hyperion <<http://www.sirsi.es/productos/hyperion.html>> para facilitar la creación de catálogos de DLOs, así como la gestión de los metadatos asociados a cada uno de ellos utilizando el DC para describir los recursos digitales. Pretende la integración de la base de datos Hyperion con el OPAC de Unicorn.
- Por su parte Ex-Libris, propietaria del SIGB Aleph, ha desarrollado MetaLib <<http://www.exlibris-usa.com/metalib>> que funciona como una plataforma óptima para administrar bibliotecas híbridas actuando como puente entre bases de datos locales y remotas. Su sistema de almacenamiento es una base de bases de datos (*AlephData Store*) que almacena recursos heterogéneos tanto en sintaxis como en estructura. Para ver el funcionamiento de este sistema: <http://www.metalib.com/V> (modo *Guest*). Más recientemente (Junio 2001) Ex-Libris presentó un módulo para la gestión de biblioteca digitales denominado DigiTooLibrary <<http://www.exlibris-usa.com/digitoolibrary>>.

cautela de que un DLO puede, y suele, cambiar desde que se genera el registro de metadatos asociado a él.

De todas las opiniones que hemos consignado aquí relativas a la asignación y almacenamiento de los metadatos se puede decir que, teniendo en cuenta el reconocimiento general del coste y la complejidad de creación de los metadatos, así como la conveniencia de asignarlos en el propio documento (CIP), la opción ideal (y más económica) es que sean los propios creadores de información los responsables también de los metadatos. No obstante es necesario contar con modelos sencillos que permitan, a los usuarios no especializados en técnicas de descripción e indización bibliográfica, introducir la metainformación a partir de herramientas o interfaces gráficas fáciles de usar. Por otra parte, si analizamos con detenimiento modelos como el de Desai atisbamos una valoración positiva a la experiencia y el conocimiento de los profesionales catalogadores y documentalistas para una mejor y más rentable (en términos de recuperación) asignación de los datos sobre los datos. Thomas y Griffin⁹³, por ejemplo, afirman que probablemente ningún sistema automatizado pueda describir información electrónica sin la asistencia humana; el hombre continuará jugando un papel importante en el proceso de recuperación de información asistida. En cuanto a la forma de almacenamiento, la tendencia, desde nuestro punto de vista, debe de ser intermedia, es decir, que los metadatos se codifiquen según un esquema en el propio documento (de tal forma que el usuario también pueda visualizarlos) y que esta misma información se extraiga de manera automática en una base de datos, ya que el procesamiento de búsqueda sobre el código completo del documento ralentizaría el sistema de recuperación.

Quizás podamos prever en todo ello, de forma subliminal, nuevas funciones para el profesional de la información que se enfrente a los retos de la nueva era de la

⁹³ Charles F. Thomas and Linda S. Griffin. Who Will Create the Metadata for Internet? [documento HTML]. *First Monday*, December 1998, vol. 3, nº 12,. Disponible en: http://www.firstmonday.dk/issues/issue3_12/thomas/index.html (consultado el 29 de enero de 2001).

información; máxime si entendemos la información Web dentro del, cada vez más importante, mercado de la información electrónica, donde la información de calidad tiene un precio y la figura del profesional que asigna metadatos dentro de un sistema concreto de información se puede estimar como un valor añadido a esa información. Tal vez se trate sólo de una redefinición del momento en que el "catalogador" —o la nueva figura del creador de metainformación digital— desempeñe su papel en la descripción cualificada en el origen o momento de gestación del documento digital. En cualquier caso, como veremos a lo largo de esta tesis, las realidades en modelos de gestión de metadatos en bibliotecas y servicios cualificados de información digital, son muy diversos.

4.3. Cómo asociar metadatos a los recursos de información electrónica: herramientas y aplicaciones

A lo largo de todo este capítulo hemos hecho alusión a la importancia de contar con herramientas sencillas que faciliten a los creadores de información Web, la tarea de codificar metadatos. Podríamos establecer incluso un principio fundamental relacionado con este tema: cuanto más fácil sea la creación de metadatos, más probabilidad habrá de que se asignen y más eficiente será el proceso. Por ello, trataremos ahora de presentar la realidad de las herramientas existentes en este sentido, diferenciando su utilización y nivel de detalle, aunque sin recomendar ninguna en especial, ya que la elección de un software para la creación de metadatos, dependerá de la finalidad de éstos (aumentar la relevancia de un motor de búsqueda determinado o ser el soporte de un sistema documental de recuperación de información, entre otras). En cualquier caso, trataremos a continuación el último aspecto que anunciamos con relación a la asignación de metadatos, el *cómo*.

Antes de abordar este tema, volvemos a plantearnos cuestiones como estas: si la finalidad de los metadatos es mejorar la recuperación de los buscadores genéricos, o ¿realmente es necesario que todos los documentos Web lleven aparejada su metainformación normalizada y estructurada? ¿o sólo deben consignarla los recursos

de calidad? ¿qué se entiende por recursos de calidad? ¿quién selecciona la calidad de la Red? A pesar de que ninguna de estas cuestiones es baladí para determinar cuándo y en qué casos, vinculados a qué tipo de información, y con qué herramientas se deben crear metadatos, como anticipamos en la introducción de este capítulo, y en la metodología de este trabajo (2.2) tratar el tema de la evaluación-calidad de recursos Web trasciende la finalidad de esta investigación. Por ello, nuestra tesis será apoyar la utilización de metadatos para recursos de información que formen parte de entornos finitos, abarcables y susceptibles de planificación, que respondan a una suerte de biblioteca digital —como por ejemplo haremos en el capítulo 9— pero sin establecer criterios sistematizados y universales para determinar su aplicación en relación con la calidad de los recursos. Ello no obsta para que también consideremos el uso de los metadatos destinados a los motores de búsqueda generales (capítulo 7), ni para que analicemos ahora las plantillas y programas que existen para aplicar metadatos con esta intención.

Es evidente que cualquiera puede publicar su página Web en Internet ya que el lenguaje HTML es sencillo y bastante indulgente con los errores formales⁹⁴, además los editores WYSIWYG, cada vez más abundantes y perfectos, facilitan en cierta medida esta tarea. No obstante, los autores no especializados rara vez consignan datos sobre los datos siguiendo un modelo o estándar concreto ni una descripción que no sea, como mucho, un conjunto de palabras en lenguaje natural y en su propio idioma, normalmente, completando las etiquetas, `<META NAME="keyword">` y

⁹⁴ Con esto nos referimos a que no es necesario que un documento siga a la perfección la semántica y sintaxis de HTML para que se pueda publicar. De esta forma se publican documentos sin cabecera `<HEAD>` y, más aún, sin metadatos. Por ello, en los motores que permiten realizar búsquedas por campos (*v.gr.* Altavista, y las opciones de búsqueda avanzada de casi todos los sistemas de recuperación de la WWW) no aprovechan esta posibilidad, bien por la negligencia de los autores de los documentos que no utilizan las metaetiquetas HTML, bien por abuso de éstas (*spamming. Vid. infr.* y *Cfr.* Glosario). No es raro por ello que en los resultados de los motores de búsqueda se contemplen documentos encabezado por "no title", aunque en el cuerpo del documento incluyan la palabra clave de búsqueda. *Cfr.* 5.1.2.

<META NAME="description">, que son las que, mayoritariamente, contemplan los motores de búsqueda generales para indizar las páginas Web y para calcular su relevancia basándose en las palabras de búsqueda.

La mayoría de las herramientas que surgen en Internet con relación a la asignación de metadatos, son meros formularios Web que, una vez completados, permitirán a los usuarios insertar las etiquetas <META> en la cabecera de un documento HTML y aumentar así, las posibilidades de aparecer en los primeros puestos de resultados de los sistemas de búsqueda generales de Internet. La proliferación de estas herramientas o plantillas, así como la falta de rigor por parte de los autores a la hora de describir el contenido de sus páginas (*spamming*⁹⁵), responde al fenómeno que Weinheimer denomina "*las guerras de metadatos*" que no consisten en otra cosa que en *un desesperado desorden por estar entre los diez primeros resultados de un motor de búsqueda*⁹⁶. Estas guerras de metadatos que responden a la

⁹⁵ Vid. Glosario.

⁹⁶ James L. Weinheimer. *Op. cit.*, p. 20. Ese "desorden" al que se refiere este autor va más allá de la mera proliferación de plantillas o herramientas para asignar metaetiquetas en aras a aumentar la visibilidad y la preeminencia en los resultados de los motores de búsqueda. Cada vez más, abundan en la Red empresas y sedes Web dedicadas exclusivamente a aumentar el tráfico de los sitios Web, como por ejemplo:

- BizWeb: <http://www.bizweb2000.com>
- Position solutions <http://www.positionsolutions.com>
- Promoteweb: <http://www.promotewebs.com>
- Spider Food: <http://www.spider-food.net>
- Web Site Promotion Tools: <http://www.websitepromosoftware.com>
- Wordtracker <<http://www.wordtracker.com>>, donde tras un sencillo registro, nombre y dirección de correo electrónico, te envían, una vez por semana, un listado de las 500 palabras claves más buscadas en la Web. top500report@wordtracker.com. *Eva, your first Wordtracker keyword report!* [correo electrónico personal] 14 de abril de 2001; 23:28. Disponible en: emendez@bib.uc3m.es. En este listado, las palabras clave más buscadas en Internet en la segunda semana de abril del 2001 fueron: mp3 (34.660) y games (22.470) seguidas de hotmail, yahoo, cars y music (se había excluido previamente en la petición, las *keywords* que evocaran pornografía). Además de este servicio gratuito, Wordtracker ofrece la posibilidad de comprar <<http://www.wordtracker.com/topwords/index.html>> listados de hasta 500.000 palabras clave al precio de £1370.

obsesión por ganar relevancia en los sistemas de búsqueda genéricos, nos permiten hacer una primera gran clasificación de las herramientas para la asignación de metainformación: 1) las destinadas a la creación de metaetiquetas y 2) las herramientas destinadas a la creación de metadatos⁹⁷.

La realidad de las aplicaciones para la producción de metainformación, en la que se funda esta división, corrobora la diferenciación que marcamos a lo largo de este capítulo, y en general de esta tesis doctoral: una cosa es la pretensión de mejorar la búsqueda genérica en Internet a través del establecimiento y aplicación de estándares de metainformación, o más bien de metaetiquetas de propósito general en HTML, y otra muy distinta, la construcción de sistemas de información de calidad basados en la organización y estructuración de la información a través de metadatos. Diferenciamos ahora esta casuística a través de las herramientas.

4.3.1. Herramientas destinadas a la creación de metaetiquetas

Son aquellas herramientas o aplicaciones que tienen como finalidad generar metainformación sencilla, de propósito general, sin seguir ningún esquema o modelo específico, almacenada en HTML, que sirvan para incrementar la relevancia en los

⁹⁷ Aunque en general, no hemos dado demasiada importancia a la diferencia entre metaetiquetas y metadatos, pudiendo entenderlas incluso como sinónimos, en este punto es fundamental hacer una distinción. Utilizamos el término metaetiquetas para designar las etiquetas <META> del HTML más genéricas —fundamentalmente, <META NAME="keywords"> y <META NAME="description">— y, metadatos propiamente dichos, para designar estas mismas etiquetas cuando su atributo NAME tiene un valor específico que responde a un modelo de metadatos como por ejemplo el DC o más aún, cuando los metadatos responden a sintaxis específica de metainformación XML/RDF.

Para esta distinción podemos basarnos en la afirmación de que todos los metadatos de una página HTML son metaetiquetas, mientras que no todas las metaetiquetas son metadatos (por ejemplo la etiqueta <META> con el atributo `http-equiv`, aunque es una información sobre la información ya que establece una comunicación entre la página Web y el servidor http, no es una información sobre el contenido de la página). En la exposición arriba se matiza más precisamente esta distinción al tipificar las herramientas y aplicaciones para generar metainformación. *Vid.* Glosario, donde se incide en esta diferenciación bajo la voz *metatag*.

motores de búsqueda o para lo que, cada vez más, se denomina *promoción del/de la Web*⁹⁸. Dentro de este primer grupo podemos distinguir: plantillas Web y software cliente.

4.3.1.1 Plantillas

Las plantillas (*templates*⁹⁹) no son más que formularios Web, realizados con un lenguaje de *script*, normalmente JavaScript, con unos campos muy sencillos para que el usuario rellene, con los valores que considere oportunos, aspectos relacionados con el contenido de su página, y automáticamente se generen las etiquetas <META> que posteriormente podrá colocar en la cabecera del código fuente de su página Web.

Son muchos las *templates* con esta finalidad, normalmente divulgadas bajo eslogan como: *haga de las metaetiquetas su arma secreta*¹⁰⁰ o, con afirmaciones tan contundentes como esta: *Conseguir los máximos resultados dentro de los motores de*

⁹⁸ Kelly analiza, en un artículo publicado en Ariadne, todas las alternativas para promocionar un sitio Web que expuso en 1999 en la conferencia europea titulada *Consolidating The European Library Space*. Incide en que las técnicas para aumentar el tráfico de un recurso Web deben aplicarse antes de publicar o lanzar el sitio a la WWW, destacando, entre otras, la asignación de metadatos para la recuperación de recursos generales, lo que aquí hemos denominado, con un efecto diferenciador, asignación de metaetiquetas. Brian Kelly. Web Focus: Using the Web to Promote Your Web Site [documento HTML]. *Ariadne*, 8-Dec-1999, issue 22. Disponible en: <http://www.ariadne.ac.uk/issue22/web-focus> (consultado el 5 de julio de 2001). A pesar de las afirmaciones de Kelly y de que la concepción más popularizada de los metadatos se funda en la posibilidad de aumentar la calidad de los resultados de la búsqueda general en la WWW, como demostraremos a lo largo de esta investigación (*Vid.* Capítulo 7), las técnicas de promoción de la Web a través de metadatos han quedado mediatizadas por la propia evolución de los motores de búsqueda que excluyen expresamente, en muchos casos, las metaetiquetas en el cálculo de la relevancia de resultados.

⁹⁹ *Vid.* Glosario.

¹⁰⁰ *Make META tags your secret weapon! Only 25% of sites use them, but most search engines evaluate them when they rank your site.* MetaMechanic <<http://www.netmechanic.com/cobrand/SiteTools/meta.htm>> por ejemplo, presenta así su plantilla editora de metadatos, aludiendo a que sólo un 25% de las páginas utilizan estas etiquetas, y sin embargo la mayoría de los motores de búsqueda las utilizan para clasificar la relevancia de las páginas Web que, como veremos, no es cierto. *Cfr.* 7.2.

búsqueda depende sólo de unas pocas líneas de texto en la cabecera de su sitio Web¹⁰¹.

El funcionamiento de estas plantillas es muy sencillo y común a todas ellas. A continuación presentamos algunos ejemplos. En todos ellos hemos introducido información relativa a esta tesis doctoral como si se tratase del recurso Web al que tratamos de asignar estas etiquetas <META>:

PLANTILLA **RESULTADOS**

http://www.technobuff.com/metabuilder/generate_meta.shtml



Fig. 4. Free MetaTag Builder



Fig. 5. Free MetaTag Builder (resultados)

<http://position-it.com/meta.htm>

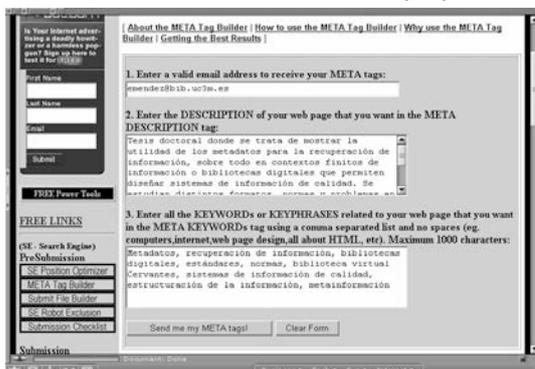


Fig. 6. MetaTag Builder Position-it

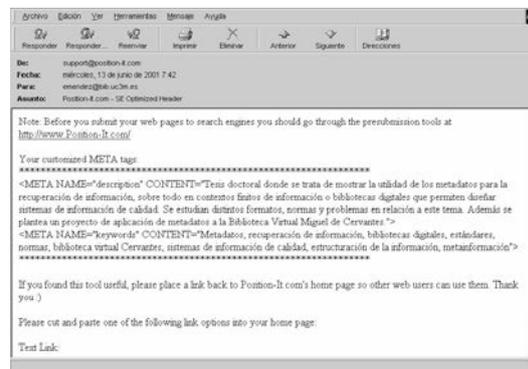


Fig. 7. MetaTag Builder Position-it (resultados)

PLANTILLA y RESULTADOS

¹⁰¹ Achieving maximum results within the search engines often depends upon just a few lines of text that go into the header of your Submit Master Web Site. SubmitMaster Free MetaTag Builder <<http://www.submitmaster.net/ezhtml/meta.html>>.

<http://www.submitmaster.net/ezhtml/meta.html>



Fig. 8. Free MetaTag Builder. Submitmaster (edición y resultados)

PLANTILLA RESULTADOS

<http://siteowner.bcentral.com/system/meta.cfm>



Fig. 9. MetaManager. Site Owner



Fig. 10. MetaManager. Site Owner (resultados)

<http://vancouver-webpages.com/META/mk-metas.html>



Fig. 11. Meta Builder 2



Fig. 12. Meta Builder 2 (resultados HTML)

Como muestran todos estos ejemplos, el funcionamiento de estas plantillas es en modo servidor y se realiza fundamentalmente en dos pasos:

- 1) El usuario introduce, a través de un sencillo formulario, el resumen y los descriptores —excepcionalmente el título en el caso de **MetaManager** <<http://siteowner.bcentral.com/system/meta.cfm>> (Figs. 9 y 10), u otra información como la dirección de correo electrónico del propietario de la página, el lenguaje, el autor, etc., en el caso de **MetaBuilder 2**¹⁰² <<http://vancouver-webpages.com/META/mk-metas.html>> (Figs. 11 y 12)— que supuestamente mejor describen el contenido del recurso Web en cuestión (en este caso, esta tesis, como si fuese un documento para la WWW)—.
- 2) El sistema genera automáticamente las etiquetas <META> en HTML y, las presenta en pantalla u ofrece la posibilidad de enviar el código a una dirección de correo electrónico que el usuario establezca, de forma exclusiva, como en el caso de **Meta Tag Builder** en el sitio de Position-it <<http://position-it.com/meta.htm>> o de forma alternativa a la presentación en pantalla, como ocurre en **MetaManager Site Owner** <<http://siteowner.bcentral.com/system/meta.cfm>>

Además de estos formularios que hemos consignado aquí a modo de ejemplo, existen otras muchas plantillas con funciones similares en torno a la edición de metaetiquetas, que si bien son metadatos de propósito general, su utilidad dista mucho de servir para fundamentar la búsqueda y recuperación de información de calidad en Internet.

¹⁰² Hasta noviembre de 1999, Vancouver Webpages también mantenían una plantilla para la creación de metadatos siguiendo el modelo Dublin Core <<http://vancouver-webpages.com/META/mk-dublin.html>>. Sin embargo, en la actualidad sólo mantienen MetaBuilder 2 para el diseño de metaetiquetas de carácter más general.

4.3.1.2. Software cliente

Dentro de las herramientas destinadas a la asignación de metaetiquetas para mejorar o incrementar las posibilidades de aparecer en un motor de búsqueda de carácter general, también tenemos que hablar de distintas aplicaciones cliente que se pueden obtener, en la mayoría de los casos *free/shareware*, con esta finalidad y que han pasado a formar parte de una categoría, cada vez más importante, en los directorios de software, denominada *promoción de la Web*¹⁰³.

A continuación presentamos y evaluamos algunas de ellas pues, en este caso, a diferencia de lo que ocurría con las plantillas, la casuística es mucho más variada y heterogénea. Para evaluar estos software hemos utilizado el mismo supuesto de tomar esta tesis como si fuese un documento para la Web, en el caso de aquellas herramientas cliente que tienen una filosofía de funcionamiento semejante a las plantillas que hemos analizado anteriormente; y un documento HTML de nuestro Web, almacenado en local, en caso de que para la generación de metaetiquetas partan de un recurso Web preexistente, al que queremos dotar de una cabecera de metainformación más elaborada.

¹⁰³ Así nos encontramos por ejemplo, esta categoría (*Web promotion*) u otras semejantes, en distintos directorios de software de Internet:

- Dave Central, incluye dentro de la categoría de *Authoring Web*, una subcategoría dedicada a este tipo de software (*Web Promotion*): <http://www.davecentral.com/webpro.html>
- Slaughter House. *Web Site Promotion*: <http://www.slaughterhouse.com/promo.html>
- Softonic. *Promotores Web*, donde aparece la subcategoría denominada *gestores de metatags*: http://www.softonic.com/index.phtml?modo=2&n_id_plat=1&n_id=883
- Tucows. *Web Promotion Tools*: <http://tucows.uam.es/promo95.html>
- Webattack, al igual que en Dave Central, incluye *Meta Tag Tools* dentro de la categoría *Web Authoring*: <http://www.webattack.com/shareware/webpublish/swmetatools.shtml>

En todos estos servidores se pueden bajar programas relacionados con la edición o generación de etiquetas, además de otras aplicaciones para enviar las páginas a varios motores de búsqueda a la vez, y otras funcionalidades que se supone, van a incrementar o a reactivar, lo que también se denomina "el tráfico de la Web".

9XMetaMagic¹⁰⁴ es una sencilla aplicación *freeware* en la que el autor del documento introduce los contenidos relativos al título, descripción y palabras clave que corresponderán a esas metaetiquetas (Fig. 13). La manera de visualización y almacenamiento es a través de un fichero que genera en HTML que podemos salvar, de tal forma que crea automáticamente la cabecera <HEAD> del documento con las etiquetas <META>.

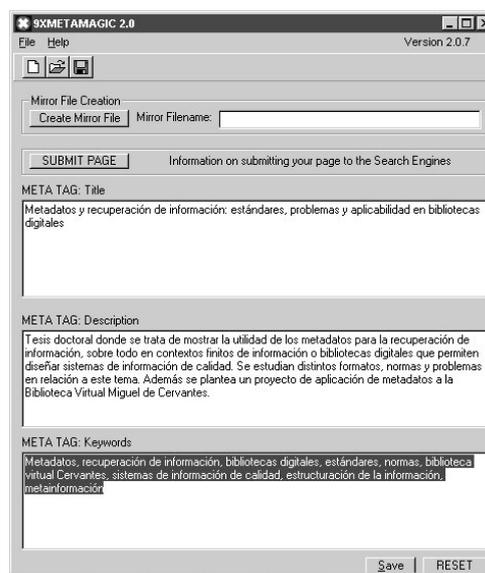


Fig. 13. 9X MetaMagic 2.07

El documento HTML guardado se visualiza con un AU (Netscape, Explorer, Opera, etc.) y aparecerá como un documento vacío en cuya cabecera estarán las metaetiquetas previamente completadas a través del formulario. Además de las etiquetas correspondientes a la metainformación que el usuario introduce, se generan dos más de forma automática (<META name="Author" content="GK Web Services"> <META name="generator" content="9xMetaMagic 2.0">) para consignar la autoría de la aplicación. Estas etiquetas (en el caso de que los buscadores las utilizasen para indizar) proporcionan una información errónea, pues el autor del recurso no es la empresa que ha desarrollado el software.

¹⁰⁴ Programa-versión: 9XmetaMagic 2.07

Homepage:<http://www.gkweb.com/9xmetamagic2.html>

Tipo: Freeware

Plataforma: Win 95/98, Win NT 3.51, Win NT 4.0

Nombre (tamaño) del fichero: 9xmm207.zip (1993 Kb)

Descarga: <ftp://ftp.gkweb.com/pub/9xmm207.zip>

Fecha de descarga: 20/03/2001 18:21

A través de la opción *submit page*, este programa dirige al usuario a una página Web con información acerca de cómo simplificar el proceso de envío del documento HTML a distintos buscadores¹⁰⁵. Esta información supone un servicio más de la empresa que ha realizado 9XmetaMagic (GK Web Services) donde, además de dar consejos o "trucos" para crear metaetiquetas que mejoren el rendimiento de los *search engines* en la búsqueda, ofrece servicios de pago para enviar una página a mil buscadores a la vez, que completan, junto con las metaetiquetas, la visión del proceso de promoción de la Web de esta empresa.

Con el mismo nivel de detalle en la descripción y con una funcionalidad muy parecida, existen otros muchos programas *free/shareware* que permiten la edición de metaetiquetas de forma semejante a 9XmetaMagic. Por ejemplo **Meta Tag Builder**¹⁰⁶ realizado por SiteUp Networks, que se autodefine como *la plataforma de lanzamiento para la promoción online*¹⁰⁷. Meta Tag Builder facilita igualmente, a través de un sencillo formulario, la generación del código de la cabecera <HEAD> de un documento HTML, creando el elemento <TITLE> y las etiquetas <META name="author"> <META name="keywords"> y <META name="description">, en cuyo atributo content, aparecerá la información que haya completado previamente el usuario. Igual que ocurría en 9XMetaMagic, crea <META name="generator"> cuyo contenido siempre será "SiteUp Internet Promotions

¹⁰⁵ Submitting your Web Page: <http://www.gkweb.com/submit/homepage.html>

¹⁰⁶ Programa-versión: Meta Tag Builder / Meta-Tag Generator
Homepage: <http://www.siteup.com/meta.html>
Tipo: Freeware
Plataforma: Win 95/98, Win NT 3.51, Win NT 4.0
Nombre (tamaño) del fichero: metabuilder12.zip (1399 Kb)
Descarga: <ftp://siteup.com/metabuilder12.zip>
Fecha de descarga: 22/03/2001 17:30

¹⁰⁷ Siteup Networks: *The launchpad for online promotions* <<http://www.siteup.com>>.

Meta-Tag Builder", sin embargo, en este caso el autor (<META name="author">) es el autor del documento.

En la misma línea de la sencillez y *freeware* podemos mencionar también: **Legacy Meta Tag Builder**¹⁰⁸, que además de la opción de copiar y pegar el código que genera tras rellenar el formulario, también permite guardarlo como documento .txt.; incluye también un manual de autoayuda, que se visualiza en el propio navegador sencillo que incluye la aplicación, y que completa con detalle las tácticas de los motores de búsqueda. Finalmente este programa, dentro de su ayuda, remite a la página Web de United Webmasters <<http://www.unitedwebmasters.com>>, donde, entre otras muchas cosas también se puede acceder a una plantilla¹⁰⁹ generadora de metaetiquetas y que funciona en modo servidor.

Similar, aunque *shareware*, es el caso de **MetaMaster**¹¹⁰. Un software realizado por Andrew Sain en Visual Basic para la edición de metaetiquetas en HTML que, una vez generadas se visualizarán en un editor de texto propio donde se podrán además, modificar y guardar tanto en .txt como en .htm. Aunque la versión

¹⁰⁸ Programa-versión: Legacy Metatag Builder 2.0

Homepage: <http://legacysoftware.cjb.net>

Tipo: Freeware

Plataforma: Win95/98/NT/2000/ME

Nombre (tamaño) del fichero: LSmeta.zip (2415 Kb)

Descarga: <http://members.nbc.com/legacyarch/programs/LSmeta.zip>

Fecha de descarga: 14/04/2001 23:34

¹⁰⁹ Para acceder a esa plantilla: <http://www.unitedwebmasters.com/meta.htm>

¹¹⁰ Programa-versión: MetaMaster 2000 v. 4.0; MetaMaster v. 5.0

Homepage: <http://www.reallybig.com/mm/>; http://www.net-matrix.com/mm_info.asp

Tipo: Shareware (\$15,00). No caduca pero es una versión reducida.

Plataforma: Win 95/98, Win NT 3.51, Win NT 4.0. Win2000

Nombre (tamaño) del fichero: meta.zip (1572 Kb) v. 4.0 meta.zip (1523 Kb) v. 5.0

Descarga: <http://www.net-matrix.com/mm/meta.zip>

Fecha de descarga: 15/10/2000 21:45 (v.4.0) 19/06/01 5:48 (v.5.0)

shareware es una versión reducida¹¹¹, a parte de las etiquetas que generaban sistemas como MetaTag Builder o 9XMetaMagic, Metamaster permite crear también las etiquetas `<META NAME="Robots">` `<META NAME="Rating">` `<META NAME="Revisit">`, aunque con unas posibilidades de selección de valores escasas. En el caso de este software, hemos tenido oportunidad de evaluar dos versiones (4.0 y 5.0), pero no hay cambios substanciales de una a otra, salvo un valor más para el contenido de la etiqueta de valoración (`rating`) y la posibilidad de informar al servidor sobre el control de la memoria caché (`Cache-Control`), en la versión superior.

En otro nivel de complejidad, aunque no mucho mayor, existen aplicaciones como **Auto Meta Tags**¹¹², *shareware* que sirve para crear las etiquetas `<META NAME="keyword">` y `<META NAME="description">`, el título, e incluso etiquetas ocultas. La diferencia con el resto de herramientas que hemos mencionado hasta ahora, es que puede funcionar de forma automática o manual. Su funcionamiento automático consiste en la generación de un título para el recurso y de la extracción de *keywords*, sin embargo, como ya hemos comentado, es mucho más fiable la asignación manual de estos contenidos. También incluye una función de búsqueda y reemplazo, que promueve el intercambio de palabras clave "flojas" por otras mejores para aparecer en los sistemas de búsqueda de la Web, adquiridas con la

¹¹¹ En realidad la única función que adquiere el programa (v. 4.0) al registrarlo es un editor de colores que produce la codificación RGB para el fondo, el texto o los enlaces de la página Web. Esto demuestra que este tipo de programas, además de para la promoción de la Web, se conciben como ayuda al diseño de páginas HTML.

¹¹² Programa-versión: Auto Meta Tag 1.02.01

Homepage: http://members.nbc.com/go_software/products.html

Tipo: Shareware (\$24,95). Caduca a los 15 días o a los 25 usos de la aplicación. Protegido contra *cracks*.

Plataforma: Win 95/98, Win NT 3.51, Win NT 4.0. Win2000

Nombre (tamaño) del fichero: autometa.zip (486 Kb)

Descarga: <http://www.websitepromotionsoftware.com/autometa.zip>

Fecha de descarga: 22/03/2001 21:41

función *Scan*. Lleva un navegador Web incorporado donde se puede previsualizar el documento HTML en el que se está trabajando (Fig. 15) y una función de compresión del archivo HTML, eliminando líneas en blanco y espacios, de tal forma que se aligeren los archivos y sean más rápidos de descargar.

Para probar Auto Meta Tags, no podemos partir del supuesto que esta tesis doctoral fuese una página publicada en la Web, ya que esta aplicación requiere partir de un fichero HTML en local al que se le asignarán metaetiquetas para aumentar su visibilidad ante los motores de búsqueda una vez publicado. Esta característica demuestra que dicho software concibe el proceso de asignación de etiquetas <META> como parte del proceso de diseño y creación de contenidos HTML, esto es, la visión más elemental de metadatos de autor/editor de recursos generales Web. Por ello, hemos seleccionado un documento .htm de nuestro Web personal que tenemos almacenado en local y que se trata de una publicación fundamentalmente textual que lleva por título, *RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio*¹¹³.

En primer lugar, trabajamos con Auto Meta Tags de forma automática (Fig. 14), que nos da la posibilidad de validar los títulos que propone sucesivamente: 1) *Helvetica modelo de datos sintaxis y esquema RDF Helvetica* 2) *Helvetica bibliografía* 3) *De modelos de metadatos surgidos en los últimos ños*, 4) *Y cubrir las necesidades de los vendedores y proveedores de información* 5) *Los modelos de metadatos con los modelos entidad relación vid Tim Berners Lee*. Obviamente, como ninguno de estos títulos es adecuado y el sistema no nos ofrece más que estas cinco opciones sucesivas,

¹¹³ La versión publicada de este recurso está disponible en: <http://www.bib.uc3m.es/~mendez/publicaciones/7jc/rdf.htm>, pero hemos trabajado con una versión en local a la que le hemos quitado toda referencia de metadatos (RDF embebido en HTML) y el contenido del <HEAD> de la versión publicada en la Web de este documento, para que fuese Auto Meta Tag, el que generase estos datos y metaetiquetas.

elegimos cualquiera de ellos para seguir adelante. No obstante, si previamente rellenamos la etiqueta <TITLE> de la cabecera, manualmente y en una sola línea, reconoce el título que le hemos asignado y lo pone como primera opción para validar.



Fig. 14. Auto Meta Tags (automático)



Fig. 15. Auto Meta Tags (manual)

El siguiente proceso que realiza el sistema de forma automática, es la creación de una descripción, donde nos ofrece una mayor cantidad de alternativas que para el título, pero igual de peregrinas e inadecuadas, mezclando valores de los atributos HTML, o las entidades de carácter, como los acentos (p. ej., ó=ó), con el contenido propiamente dicho. Al mismo tiempo va generando los descriptores, y finalmente crea las metaetiquetas, siendo el resultado final, este:

```
<!doctype html public "-//w3c//dtd html 4.0 transitional//en">
<html>
<head>
<title>!!! UN MODELO DE METADATOS FLEXIBLE PARA LAS BIBLIOTECAS
DIGITALES DEL PR OACUTE XIMO !!!</title>
<meta name="description" content=" digitales y o virtuales si bien
es cierto que existe una tendencia incipiente">
<meta name="keywords"
content="xml,web, volver ,una,uacute,sup,rdf, que,por, para,org,oacute,m
odelo,metadatos, los, las, justify, informaci, iacute, helvetica, eacute, di
v, del, como, aacute, 999, aacute 999">
<!-- un modelo de metadatos flexible para las bibliotecas digitales
del pr oacute ximo and even
xml,web, volver ,una,uacute,sup,rdf, que,por, para,org,oacute,modelo,met
adatos, los, las, justify, informaci, iacute, helvetica, eacute, div, del, com
o, aacute, 999 Auto Meta V.01.01 --> </head>
```

Al igual que demostramos en el apartado anterior al hablar de Klarity, el resultado indica que no podemos confiar la asignación del contenido de los metadatos o *metatags* a un sistema automático. Si utilizamos Auto Meta Tags de forma manual (Fig.15), donde la aplicación nos deja seleccionar los descriptores de entre los que el sistema propone automáticamente, los resultados no mejoran demasiado. El proceso final, tanto si utilizamos el modo automático como el manual es el mismo, el software permite guardar los cambios hechos en el documento HTML que no afectarán a su presentación, pero completarán la cabecera de metainformación. Si bien es cierto que con documentos HTML en inglés este sistema es algo más preciso en la indización automática que realiza, la creación de metaetiquetas con este software es, aunque sencilla, insuficiente, incluso teniendo en cuenta que la supuesta finalidad de estas *metatags* es mejorar el tráfico de la Web sobre las páginas.

MetaDemon¹¹⁴ es otra herramienta, también *shareware* para la edición de etiquetas <META> similar al último software que hemos descrito, en el sentido de que trabaja sobre un fichero HTML existente en local, y que también presenta la posibilidad de extracción automática de palabras clave con opción de validar su presencia en las metaetiquetas. La característica más relevante de MetaDemon, frente a otros software, es la gran importancia que otorga a la utilidad de los metadatos para evaluar o clasificar los contenidos en aras a la protección infantil ante la Red. Para probar esta herramienta, hemos utilizado el mismo ejemplo que en el caso anterior, la página Web *rdf.htm* almacenada en local. Una vez que abrimos el documento lo primero que hace este programa es chequear si existe la etiqueta <TITLE> e

¹¹⁴ Programa-versión: MetaDemon. 2.1.1.

Homepage: <http://www.rd-soft.com/downloads/metademon.html>

Tipo: Shareware (\$10)

Plataforma: Win 95/98

Nombre (tamaño) del fichero: metademon.zip (1640 Kb)

Descarga: <http://www.rd-soft.com/files/metademon.zip>

Fecha de descarga: 06/04/2001 13:53

identificar este elemento dentro del campo correspondiente de su formulario, así como analizar automáticamente el fichero HTML para extraer los descriptores (Fig. 16).

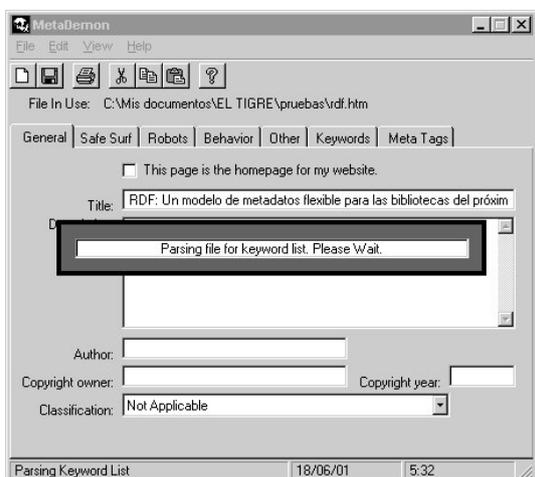


Fig. 16. MetaDemon 2.1.1.1.

Una vez que finaliza el análisis de *keywords*, nos permite rellenar distintos campos que serán el contenido de las metaetiquetas a través de distintos niveles: en primer lugar, un nivel general que incluye el título (extraído automáticamente si existe en la cabecera del documento), el resumen (*description*), el autor, el propietario y el año del copyright, y una clasificación del contenido que puede ir desde la

declaración de que se trata de una página personal, hasta la matización del tipo de contenido que representa la página Web (información para adultos, sobre Internet, sobre viajes, gubernamental, ocio, etc.); en un segundo nivel de descripción (*Safe surf*¹¹⁵) admite consignar información más precisa sobre la valoración del contenido, definiendo el rango de edad y si se trata de información sobre sexo, violencia, intolerancia, etc. y en qué grado se presenta ésta; en un tercer nivel de metainformación, se expresará toda la información relativa al comportamiento de los robots con relación a la página; en el cuarto nivel (*behaviour*) se definen lo que

¹¹⁵ Esta opción se denomina *Safe Surf*, porque la clasificación de contenidos que permite responde al estándar de ese nombre, diseñado en 1995 por y para los padres de forma que cada familia pueda tomar decisiones informadas sobre la accesibilidad de los contenidos *online*. *Safe Surf* <<http://www.classify.org/safesurf>> dispone también de un *plug-in* <<http://www.classify.org/safesurf/iesetup>> que permite configurar el navegador Internet Explorer para categorizar los contenidos según este estándar. Como mencionamos al hablar de la utilidad de los metadatos para la valoración previa del tipo de información existen diversas normas e iniciativas con esta finalidad, entre las que destaca el estándar del W3C, PICS. *Cfr.* Capítulo 3, nota 80.

algunos editores de páginas Web denominan "metavariables" del sistema, en las que se cumple el pequeño matiz diferencial que hemos señalado de ser metaetiquetas, pero no metadatos, tal es el caso p. ej. de la etiqueta <META CONTENT=http-equiv>; dentro del nivel de descripción denominado en este software *Others* se puede seleccionar la codificación de caracteres, el lenguaje de *scripts* utilizado, así como el idioma o variedad idiomática del documento; en la opción de *keywords* presenta un listado alfabético con todos los descriptores del recurso que estamos describiendo, donde el usuario podrá seleccionar aquellas palabras clave extraídas que no quiere que aparezcan en sus metadatos, así como la forma de presentación. En el proceso final del funcionamiento de MetaDemon se generan las metaetiquetas, dando como resultado de todo el nivel de detalle de la descripción, el siguiente código fuente, que se puede copiar y pegar o bien guardarlo como formato txt:

```
<HEAD>
<META HTTP-EQUIV="Content-Language" CONTENT="es-ES">
<META NAME="revisit-after" CONTENT="15 days">
<META NAME="robots" CONTENT="Index,Noimageindex,Imageclick,Follow">
<META HTTP-EQUIV="PICS-Label" CONTENT='(PICS-1.0
"http://www.classify.org/safesurf/" 1 on "2001.06.16T07:04-0100 r
(SS~000 1 SS~001 3)')'>
<META NAME="distribution" CONTENT="Local">
<META NAME="classification" CONTENT="INTERNET:Resources">
<META NAME="copyright" CONTENT="Copyright 1999 by Eva Méndez">
<META NAME="author" CONTENT="Eva Méndez">
<META HTTP-EQUIV="expires" CONTENT="mar, 18 sep 2001 06:50:33 GMT">
<META NAME="description" CONTENT="Esta comunicación versa sobre el
Resource Description Framework (RDF), un modelo de metadatos basado
en XML, que promete ser un estándar flexible para la estructuración
de la información de las bibliotecas digitales del siglo XXI. Se
describen el contexto y las características principales de dicho
formato, así como sus posibilidades y potencial desarrollo para
optimizar la recuperación de información en Internet.">
<META NAME="keywords"
CONTENT="accesible,biblioteca,digital,interoperabilidad,marcados,met
adatas,metadatos,metalenguajes,normalizaci,rfd,webs">
<TITLE> RDF: Un modelo de metadatos flexible para las bibliotecas
del próximo milenio </TITLE>
</HEAD>
```

Como muestra el código fuente, MetaDemon permite generar una cabecera de metaetiquetas mucho más completa que en las otras aplicaciones que hemos mencionado hasta ahora, con la ventaja de que también se pueden crear metadatos basados en el estándar PICS (*Platform for Internet Content Selection*) y en el código

de *Safe Surf* que sirven para evaluar el contenido del recurso y preservarlo del acceso a menores.

Color and Metatag Buster¹¹⁶ es otra aplicación semejante, también *shareware*, que mezcla utilidades de edición de color en páginas Web, con la creación de etiquetas <META> y con un generador de señalización para "engañar" a los motores de búsqueda de tal forma que las páginas Web tengan mayor y mejor presencia en ellos. En cuanto a editor de metaetiquetas, este software es muy sencillo (Fig. 17), consta de un formulario que permite rellenar los campos de: palabras clave, descripción, autor y copyright, así como información específica para los buscadores (la frecuencia con la que queremos que vuelvan a revisar la página —<META NAME="revisit-after" CONTENT="X weeks">—, o el tipo de robots que queremos que concurren para indizarla).

Nuevamente utilizamos el supuesto de esta tesis doctoral como si se tratase de un documento Web que queremos publicar en Internet, ya que Color and Metatag Buster supone la creación de metainformación de forma independiente al fichero HTML y éste puede ser un documento publicado ya en la Web. Una vez completada la información que implican los campos que incluye, se generan las metaetiquetas correspondientes, que se copian en el portapapeles de Windows, de tal manera que podamos pegar el código resultante en la cabecera <HEAD> de nuestro documento HTML.

¹¹⁶ Programa-versión: Color and Metatag Buster. 2.02/2
Homepage: <http://www.tashcom.com/colorbuster.html>
Tipo: Shareware (\$15,00).
Plataforma: Win 95/98, Win NT 3.51, Win NT 4.0
Nombre (tamaño) del fichero: cbuster.zip (1122 Kb)
Descarga: <http://www.tashcom.com/cgi-local/load.pl?../files/cbuster.zip>
Fecha de descarga: 22/03/2001 19:54

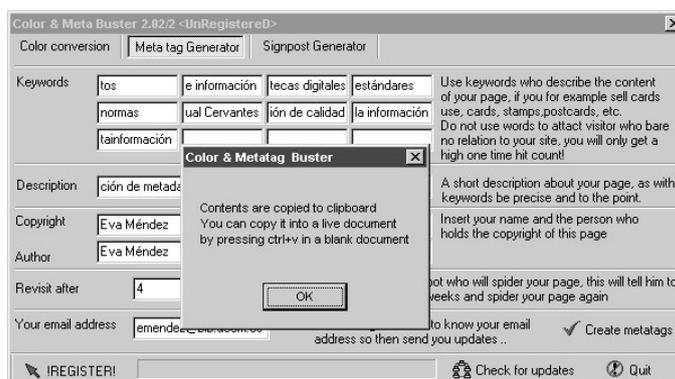


Fig. 17. Color and Metatag Buster 2.02

El código que genera MetaTag Buster para nuestro ejemplo es el siguiente:

```
<-- generated by color & metatag buster (C) Tashcom software
(http://www.tashcom.com) -->
<-- start of meta-tags -->
<link REV="made" href="mailto:emendez@bib.uc3m.es">
<meta NAME="keywords" CONTENT="Metadatos, recuperación de
información, bibliotecas digitales, estándares, normas, Biblioteca
virtual Cervantes, sistemas de información de calidad,
estructuración de la información, metainformación,">
<meta NAME="description" CONTENT="Tesis doctoral donde se trata de
mostrar la utilidad de los metadatos para la recuperación de
información, sobre todo en contextos finitos de información o
bibliotecas digitales que permiten diseñar sistemas de información
de calidad. Se estudian distintos formatos, normas y problemas en
relación a este tema. Además se plantea un proyecto de aplicación de
metadatos a la Biblioteca Virtual Miguel de Cervantes.">
<meta NAME="copyright" CONTENT="Eva Méndez">
<meta NAME="author" CONTENT="Eva Méndez">
<meta NAME="revisit-after" CONTENT="4 Weeks">
<meta NAME="Robot" CONTENT="ALL">
<meta NAME="rating" CONTENT="General">
<meta NAME="DC.Title" CONTENT="Color & Metatag buster 2.02/2">
<meta NAME="DC.Creator" CONTENT="Color & Metatag buster 2.02/2">
<meta NAME="DC.Description" CONTENT="keyword & content builder --
http://www.tashcom.com">
<-- end of meta-tags -->
```

Lo que más llama la atención frente a los software cliente que hemos analizado hasta el momento, es la utilización, al final de tres etiquetas en formato DC (DC.Title, DC.Creator y DC.Description) con contenidos únicamente alusivos al propio software. Esto revela, por un lado el reconocimiento tímido del DC como modelo de metadatos general para la recuperación de información en la Web,

pero por otra parte, demuestra el poco reconocimiento de este formato que tienen los sistemas de búsqueda, al no utilizar el estándar del DC para todo el código de metaetiquetas¹¹⁷.

Finalmente, vamos a describir otras herramientas que, a pesar de trabajar únicamente con etiquetas <META> en HTML, son algo más que meros editores de metaetiquetas, tal es el caso de Meta Builder, MetaFormer, MetaMaximicer, Metty y BHead.

La primera de ellas, **Meta Builder**¹¹⁸, es también una herramienta *shareware* poco conocida pero muy potente pues se trata no sólo de un editor-generador de metaetiquetas, sino que también puede funcionar como un validador de éstas además de que permite ir construyendo una base de datos con todas las metaetiquetas utilizadas en lo que, en el seno de este software, se denominan "proyectos". Como la mayoría de estas aplicaciones informáticas que utilizan las etiquetas <META> como una táctica más para incrementar el tráfico de la WWW, da una gran importancia a las palabras clave. Además de que el usuario puede describir el documento con sus

¹¹⁷ De todos los *share/freeware* que hemos analizado en esta sección, los únicos que soporta las etiquetas del modelo Dublin Core —si excluimos MetaBuilder que permite crear etiquetas *ad hoc*—, en formato de almacenamiento HTML, son Metty <<http://www.clickfire.com/freeware/metty>> y PromotionPro <<http://www.gammadyne.com/metamax.htm>>, que las contemplan como "otras metaetiquetas" que se pueden incluir en la cabecera de un documento HTML para aumentar su visibilidad en los sistemas de búsqueda genéricos de la WWW.

Para consultar un listado de todas las metaetiquetas posibles en HTML, tanto pertenecientes al atributo HTTP-EQUIV, como al atributo NAME, Vid. *A Dictionary of HTML Metatags* [documento HTML]. Vancouver Webpages, rev. 16 de marzo de 2001. Disponible en: <http://vancouver-webpages.com/META> (consultado el 4 de abril de 2001).

¹¹⁸ Programa-versión: Meta Builder 6.1.2

Homepage: <http://www.triserwebs.com/products/metabld.htm>

Tipo: Shareware (\$ 29,50)

Plataforma: Win 95/98, Win NT

Nombre (tamaño) del fichero: MetaBuild60.exe (3496 Kb)

Descarga: <http://www.triserwebs.com/Download/MetaBuild60.exe> (registro de nombre, *email* y país)

Fecha de descarga: 09/03/2001 20:19

propias *keywords* también tiene un generador automático de descriptores a partir del documento HTML para el que se asignarán las metaetiquetas.

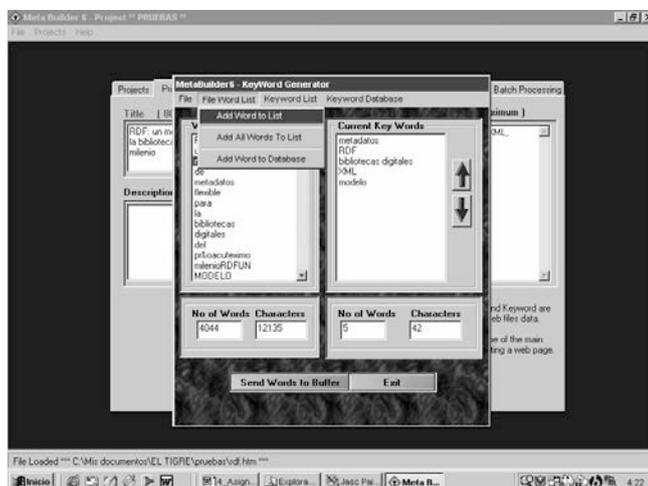


Fig. 18. Meta Builder 6.1.2.

Para analizar esta aplicación hemos utilizado la publicación sobre RDF almacenada en local (fichero *rdf.htm*) a la que venimos aludiendo como metodología alternativa para realizar la evaluación del software. Automáticamente selecciona el contenido del <TITLE> que era la única etiqueta que dejamos en la cabecera de este fichero de pruebas. Posteriormente rellenamos las etiquetas primarias (a parte del título, según este software, resumen y palabras clave). Para la asignación de descriptores utilizamos la función *generate keywords* (Fig. 18), de tal forma que el propio sistema extrae palabras¹¹⁹ del documento *rdf.htm* que podemos validarlas

¹¹⁹ Esta forma de extracción de palabras clave es una metodología antigua en indización automatizada. Como hemos dicho en otros trabajos, la extracción de palabras basada en modelos estadísticos o probabilísticos constituye la primera generación de la indización automatizada. *Vid.* José Antonio Moreiro González, Eva M^a Méndez Rodríguez. *Op. cit.*, p. 15-17.

Si bien es cierto, como muestran los resultados de la extracción automática que hace este programa, no es muy fiable (al menos con documentos en español). Esta función de MetaBuilder podría servir sin embargo, para extraer términos de un conjunto de documentos HTML que formen parte de una colección digital y utilizarse como términos candidatos de un tesoro que trabaje como herramienta de recuperación por materias de una biblioteca digital, aprovechando la función que tiene este software de ir creando una base de datos de *keywords*.

como palabras clave de la etiqueta <META NAME="keywords">. En general, como hemos visto en otras aplicaciones semejantes —como Auto Meta Tag o MetaDemon— esta función es bastante común en casi toda esta gama de software editor de metaetiquetas¹²⁰, lo realmente interesante de MetaBuilder es que permite almacenar las palabras clave en una base de datos. También completamos las etiquetas secundarias (author, copyright, language, classification, generator, rating, robots, revisit-after), incluso podemos añadir etiquetas nuevas (por ejemplo, podríamos crear etiquetas con formato DC). Una vez que hemos completado el proceso de asignación guiada de metaetiquetas, podemos almacenar éstas en una base de datos (que en este caso hemos denominado "rdf") de tal forma que podamos utilizar esas mismas etiquetas para la creación de metainformación de páginas semejantes.

Dentro de las aplicaciones cliente que estamos evaluando, que además de ser editores de *metatags* tienen otras funciones complementarias con relación a la promoción de la Web, podemos hablar también de otro sencillo, pero interesante software, **Metaformer**¹²¹. En cuanto a la edición de metaetiquetas, funciona de forma parecida a las herramientas que ya hemos comentado. La mayor particularidad que presenta es que se pueden crear tanto para ficheros almacenados en local, como

¹²⁰ También incluso en el caso de algunas herramientas gratuitas como MetaWizard:

Programa-versión: MetaWizard 1.2a

Homepage: <http://tafweb.hypermart.net/metawizard.html>

Tipo: Freeware (registro opcional)

Plataforma: Win 95/98, Win NT.

Nombre (tamaño) del fichero: mwsetup.exe (603 Kb)

Descarga: <http://www.tafweb.force9.co.uk/download/mwsetup.exe>

Fecha de descarga: 22/03/2001 21:31

¹²¹ Programa-versión: Metaformer 2.0.0

Homepage: <http://pmpo.hypermart.net/pmpo/metaformer/index.htm>

Tipo: Shareware (\$17,50); caduca a los 30 días

Plataforma: Win 95/98, Win NT 3.51, Win NT 4.0

Nombre (tamaño) del fichero: metaf.zip (1164 Kb)

Descarga: <http://pmpo.hypermart.net/pmpo/metaformer/ metaf.zip>

Fecha de descarga: 20/03/2001 17:59

como la nuestra. El código fuente que genera se visualiza en todo momento en un editor de texto, y los resultados se guardan en un archivo .htm (por defecto, tmp.htm).

Metaformer, incluye una aplicación denominada PMPro NetAnt 2000 que sirve para enviar la página Web en la que estemos trabajando, a los principales motores de búsqueda de Internet seleccionando hasta quince de ellos (Excite, Hotbot, Google, Altavista, Northernlight, etc.); a pesar de estar integrada con las funciones de asignación de metaetiquetas puede utilizarse de forma individual obviando el proceso de creación de metaetiquetas para incluirlas en el código fuente de los documentos HTML. Otra de las particularidades que hacen de este software una herramienta destacable en cuanto a su facilidad de uso, es la opción MetaFormer Wizard que ayuda al usuario a completar paso a paso todo proceso que hemos descrito de asignación de metaetiquetas y palabras clave y el envío de la página a distintos sistemas de búsqueda¹²³.

¹²³ La utilización de aplicaciones *Wizard* (que ayudan al usuario a ejecutar una tarea de forma eficaz, normalmente siguiendo la técnica de *paso a paso* o la automatización de las tareas) es cada vez más frecuente en el entorno Windows. Esta modalidad de funcionamiento es especialmente interesante en este tipo de herramientas pues, la mayor parte de los usuarios no utilizan metainformación en sus páginas por desconocimiento. Este modelo, *Wizard* o paso a paso, es el que sigue, en este caso de forma exclusiva, la aplicación Metatag Toolkit:

Programa-versión: Metatag Toolkit 1.0
Homepage: <http://www.dittotech.com/Products/MetaTag>
Tipo: Freeware
Plataforma: Windows 95, 98, NT, 2000
Nombre (tamaño) del fichero: mtsetup.exe (315 Kb)
Descarga: <ftp://ftp.dittotech.com/pub/mt/mtsetup.exe>
Fecha de descarga: 20/03/2001 18:52

Metamaximicer¹²⁴, desde el 1 de marzo de 2001 (versión 4.6) llamado **Promotion-Pro**, que es un nombre más adecuado puesto que, como MetaFormer,

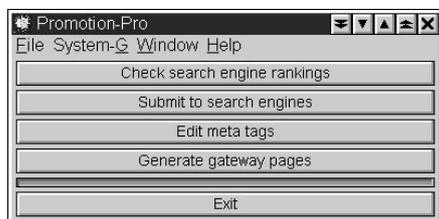


Fig. 20. Promotion-Pro 4.6 (pantalla principal)

combina la edición de metadatos con otras estrategias para la promoción de la Web (Fig. 20): comprobar el ranking de nuestras páginas Web en los once principales motores de búsqueda (*search engine ranking*), enviar nuestros documentos HTML hasta a treinta y tres buscadores distintos (*Submit to search engines*), o generar páginas pasarela (*gateway pages*¹²⁵), especie de portales de acceso que sirven para atraer mejores resultados en los sistemas de búsqueda genéricos de la WWW.

En cuanto a sus funciones como editor de metainformación, tiene una interfaz sencilla que facilita la edición de metaetiquetas en la cabecera de los documentos Web. Se pueden crear metaetiquetas tanto para un documento HTML existente almacenado en local, como para uno nuevo. Permite crear un gran número de metaetiquetas (más de 200), aunque reconoce que la mayor parte de ellas no mejorarán el *ranking* en los motores de búsqueda salvo en el título, descripción y palabras clave. Las etiquetas `<META>` que se pueden asignar con este programa

¹²⁴ Programa-versión: MetaMaximicer 3.0/ Promotion-Pro 4.6

Homepage: <http://www.gammadyne.com/metamax.htm>

Tipo: Shareware (\$ 39,00)

Plataforma: Windows 95, 98, NT, 2000, ME, XP

Nombre (tamaño) del fichero: metamax.exe / promo.exe (812 Kb)

Descarga: <http://www.gammadyne.com/promo.exe>

Fecha de descarga: 20/03/2001 18:42

¹²⁵ Estas páginas pasarela que MetaMaximicer llama *gateway pages* reciben con frecuencia también la denominación, manteniendo el mismo significado implícito: *bridge pages* o *doorway pages*. What is a Bridge or Doorway Page? [documento HTML]. Search Engine Watch, 2 de junio de 2000. Disponible en: <http://searchenginewatch.internet.com/webmasters/bridge.html> (consultado el 21 de julio de 2001). Cfr. Glosario, *gateway*.

aparecen en cinco grandes secciones (Fig. 21) donde, con una interfaz muy simple, se completa toda la información del <HEAD> relativa a:

- 1) Descripción, donde incluye los campos de título, descripción y clasificación genérica de la página, p. ej., tomando como modelo el supuesto de esta tesis como documento HTML:

```
<base href="http://www.bib.uc3m.es/~mendez">
<title>Metadatos y recuperaci&ocute;n de informaci&ocute;n:
est&acutes;ndares, problemas y aplicabilidad en bibliotecas
digitales</title>
<meta name="description" content="Tesis doctoral donde se
trata de mostrar la utilidad de los metadatos para la
recuperaci&ocute;n de informaci&ocute;n, sobre todo en
contextos finitos de informaci&ocute;n o bibliotecas
digitales que permiten dise&ntilde;ar sistemas de
informaci&ocute;n de calidad. Se estudian distintos formatos,
normas y problemas en relaci&ocute;n a este tema.
Adem&aacute;s se plantea un proyecto de aplicaci&ocute;n de
metadatos a la Biblioteca Virtual Miguel de Cervantes.">
<meta name="keywords" content="metadata,information
retrieval,digital libraries,standards,quality information
systems,information structure,metainformation126">
<meta name="classification" content="Communications and
Networking">
```

- 2) Info, donde se creará la metainformación relativa al autor, copyright, URL base, así como las etiquetas de conjunto de caracteres utilizados, idioma, etc., p. ej.:

```
<meta http-equiv="content-type"
content="text/html;charset=ISO-8859-6">
<meta http-equiv="content-language" content="es">
<meta http-equiv="pragma" content="no-cache">
```

- 3) Motores de búsqueda, donde se seleccionará el contenido para todas las metaetiquetas relacionadas con ellos, p. ej.

```
<meta name="robots" content="all">
<meta name="distribution" content="Global">
<meta name="revisit-after" content="30 days"> <meta
name="rating" content="General">
```

¹²⁶ Nótese que en este ejemplo hemos puesto los descriptores en inglés para poder aplicar la función que tiene este programa de "*phrase analysis*".

- 4) *Keywords*. En este nivel permite asignar las palabras clave de búsqueda. Aunque el sistema de introducción de las palabras es poco amigable, tiene otras funciones que compensan, como el análisis de frase y la estadística de términos utilizados (Fig. 21), o la posibilidad de visualizar un listado de palabras vacías y un listado de palabras clave para adultos.

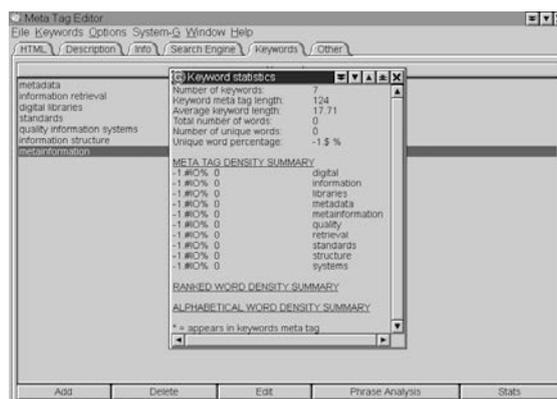


Fig. 21. Promotion-Pro 4.6 (*keywords*)

- 5) Por último, en la sección *others*, admite la asignación de hasta 190 metaetiquetas más seleccionando entre valores como, p. ej. abstract, DC.Description, MathDMV.ABSTRACT, que demuestran que este sistema ha contemplado, a su manera, la disensión semántica de los valores de las etiquetas¹²⁷; o también etiquetas relativas al PICS o las quince etiquetas básicas del Dublin Core.

Todas estas características, además de las grandes posibilidades de configuración, las utilidades añadidas que tiene (calculadora, calendario y tabla de caracteres), y la posibilidad de asociar las mismas etiquetas a varios ficheros HTML a la vez, hacen de Promotion Pro una herramienta muy potente que nos lleva a considerarla algo más que un mero editor de metaetiquetas.

¹²⁷ Cfr. 8.3.2.

Otra aplicación, que merece una mención especial, en este caso no porque tenga grandes funciones añadidas, sino porque, además de las etiquetas <META> convencionales, permite también editar metadatos DC, embebidos en la cabecera de un documento HTML, es **Metty**¹²⁸. En general, acepta la creación de hasta 33 metaetiquetas con la misma finalidad de elevar los *rankings* y la exposición global de las páginas Web. Metty puede insertar las etiquetas generadas en un fichero HTML que se haya importado bien localmente o bien de un sitio web. Este método de introducir el URL para importar la página HTML remota, al igual que en MetaFormer, sirve para chequear qué tipo de etiquetas se utilizan. Como PromotionPro, divide las etiquetas que se pueden crear con este software, en secciones, en este caso en tres: etiquetas generales, etiquetas DC, y otras. El programa también ofrece la opción de rellenar las metaetiquetas a partir de un ejemplo, en el caso de que el usuario no esté familiarizado con el significado de las etiquetas individuales. Permite generar las etiquetas y posteriormente incluirlas en el fichero HTML que estemos trabajando, de igual forma que da la opción de guardarlas individualmente o con todo el archivo HTML.

Por último **BHead**¹²⁹, un *freeware* que funciona también como un generador avanzado de metaetiquetas, al igual que los anteriores, dividiendo el trabajo en

¹²⁸ Programa-versión: Metty 1.1

Homepage: <http://www.clickfire.com/freeware/metty>

Tipo: Freeware

Plataforma: Win 95/98/NT/2000/ME

Nombre (tamaño) del fichero: metty.exe (2385 Kb)

Descarga: <http://www.clickfire.com/download/Metty.exe>

Fecha de descarga: 20/06/2001 06:27

¹²⁹ Programa-versión: Bhead 4.1.2

Homepage: <http://home5.swipnet.se/~w-52253/>

Tipo: Freeware

Plataforma: Win 95/98/NT/2000/ME

Nombre (tamaño) del fichero: bhsetup.exe (1011 Kb)

Descarga: <http://homes.swipnet.se/~w-52253/hyper/program/bhead/bhsetup.exe>

Fecha de descarga: 16/06/2001 18:52

secciones (*metatag* básicas, adicionales, donde se pueden crear etiquetas propias — incluyendo el DC o cualquier otro estándar de metadatos—, y etiquetas "extra", entendiéndose por estas últimas la función de *refresh* `<meta http-equiv="Refresh" . . .>`, y advirtiéndose que muchos buscadores no registran páginas con esta función). El valor añadido de esta herramienta es que, además puede encargarse de toda la cabecera `<HEAD>`, incluso con un editor básico de hojas de estilo CSS, que incluye el diseño de color. El resultado es un código `<HEAD>` bastante bien formado y completo que se copia en un documento para un uso posterior o para la edición, o bien guardarse en un archivo.

Finalmente, y dentro de esta categoría que estamos tratando de herramientas cliente para la creación de metaetiquetas, debemos hacer mención a un tipo especial de aplicaciones, los analizadores de etiquetas `<META>` que pueden ser de gran utilidad para realizar estudios de uso de la metainformación¹³⁰ en documentos HTML, así como para calificar su visibilidad en un motor de búsqueda.

En este sentido, debemos destacar **Meta Tag Maximizer**¹³¹, *shareware* que además de ser un editor de metadatos similar a los que hemos descrito anteriormente (establecimiento de metaetiquetas básicas, indización del texto completo a partir de la cual se pueden extraer los descriptores, etc.), tiene una función de análisis de

¹³⁰ Semejantes a los que mencionamos anteriormente (*Vid. sup.* Nota 1), o incluso de mayor profundidad. A pesar de que este tema del nivel de utilización de metadatos en páginas Web lo excluimos *ex professo* de esta tesis doctoral, consideramos que es necesario realizar muestreos de este tipo para tener una visión realista de la utilización de este método de descripción de recursos Web, sobre todo en páginas y recursos de instituciones, contenidos, etc. que pueden llegar a formar parte en un momento dado de "sistemas de información de calidad" o *subject gateway* (*Vid.* 8.2.2.)

¹³¹ Programa-versión: Meta Tag Maximizer 2.1
Homepage: <http://www.metatagmaximizer.com>
Tipo: Shareware
Plataforma: Win 95/98/NT/2000/ME
Nombre (tamaño) del fichero: metamax.exe (1421 Kb)
Descarga: <http://www.imetatag.com/metamax.exe>; <http://tucosws.uam.es/files3/metamax.exe>
Fecha de descarga: 22/06/2001 04:08

metainformación, tanto de páginas almacenadas en local, como remotas. Una vez analizadas las metaetiquetas, el sistema produce automáticamente un informe sobre ellas (Fig. 22), con relación a los siguientes ítems: etiquetas <META> encontradas en el documento, características generales de la página, un análisis de las palabras clave (que incluye tanto las encontradas en la etiqueta correspondiente, como en el cuerpo del documento), y finalmente, una serie de recomendaciones generales para aumentar el *ranking* de relevancia de dicha página en los buscadores.

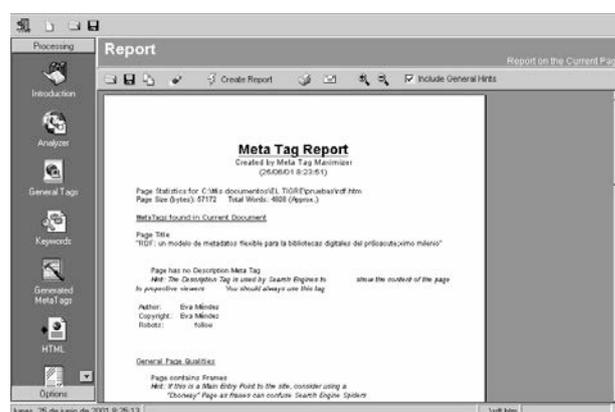


Fig. 22. Meta Tag Maximizer 2.1

4.3.2. Herramientas destinadas a la creación de metadatos

Las herramientas para la creación de metadatos, por contraposición a las anteriores (4.3.1.), son aquellas que sirven para simplificar el proceso de asignación de metadatos según un modelo o esquema (p. ej. DC, GILS, AGLS, etc.) y una sintaxis de almacenamiento particular. Aunque tácitamente el objeto de esta metainformación también podría ser algo parecido a la "promoción de la Web", en sentido estricto, podemos decir que su finalidad es la recuperación eficaz en un sistema de información de calidad en la WWW o la organización de colecciones de recursos en Internet/Intranet, en aras a la concepción de "bibliotecas digitales".

Al igual que con el software destinado a la asignación de metaetiquetas, podemos tipificar también estas herramientas en dos grandes grupos: las plantillas o *templates* que funcionan en modo servidor y los software cliente que, en muchos

casos entrañan otras funciones además de la edición del código de metadatos. Una de las clasificaciones, menos intuitiva que la nuestra, pero más aceptada es que se expone en un estudio llevado a cabo en el seno del proyecto MetaWeb del DSTC¹³² (*Distributed Systems Technology Centre*) en Australia, que distingue entre: editores y generadores, sin tener en cuenta si su forma de trabajo es un formulario en el servidor o una aplicación cliente. Los editores son aquellas herramientas que permiten la creación a través de una plantilla donde se introduce el contenido de los metadatos, posteriormente el sistema coloca dicho contenido en etiquetas <META> de HTML o en una sintaxis RDF, que finalmente se puede cortar y pegar en el código de un recurso electrónico. El segundo tipo de herramientas que distingue el DSTC, los generadores, se diferencian de los anteriores en que extraen directamente los metadatos del código HTML del documento y suelen permitir la selección del formato de codificación, HTML o XML/RDF, normalmente (a este último tipo de aplicaciones, pertenece, por ejemplo, Klarity <<http://www.klarity.com.au>> o Automatic RDF Generator <<http://scitsd.wlv.ac.uk:8080/metadata.html>>, o como explicaremos también a continuación, herramientas como DC-Dot).

Thornely¹³³ también señala dos modalidades en la creación de metadatos podemos contemplar para tipificar las herramientas a tal efecto: por un lado la codificación en tiempo real, por el propio creador del recurso Web, dentro del propio

¹³² *MetaWeb Project: Analysis of Metadata Creation Tools* [documento HTML]. Brisbane: University of Queensland, Distributed Systems Technologies Centre, rev. 31 de agosto de 1999. Dponible en: <http://www.dstc.edu.au/Research/Projects/metaweb/toolpost.html> (consultado el 18 de febrero de 2001). En el ámbito de la investigación australiano, son varios los autores que distinguen estos dos tipos de herramientas para la creación de metadatos. V. gr. se mencionan estos dos tipos de aplicaciones en el recurso creado por la Biblioteca Nacional de Australia (NLA) sobre metadatos: *Meta Matters* [documento HTML]. Canberra: National Library of Australia, 24 de mayo de 1999. Disponible en: <http://www.nla.gov.au/meta> (consultado el 17 de febrero de 2001). Thornely utiliza esta misma denominación, editores y generadores, para clasificar las aplicaciones que generan metadatos Dublin Core. Jennie Thornely. *The How of Metadata: Metadata Creation and Standards. Cataloguing Australia*, 1999, vol. 25, nº 1/4, p. 32.

¹³³ *Ibid.*, p. 33.

proceso de creación del contenido de las páginas, y *retro-codificación* realizada por otros agentes, después de que la página haya sido creada. A pesar de lo elocuente de estas clasificaciones, la primera basada en el criterio de la automatización de la extracción de metadatos, y la segunda, en el momento en que éstos se crean, hemos preferido utilizar nuestra propia tipificación de estas herramientas, por tres razones fundamentales:

- a) En primer lugar, por coherencia con la división que realizamos al hablar de las aplicaciones destinadas a la creación de metaetiquetas de carácter general que analizamos en el apartado anterior.
- b) En segundo lugar, porque la mayoría de las aplicaciones, al igual que en las destinadas a la promoción de la Web, combinan la asignación directa con la extracción automática y permiten asimismo la asignación de metaetiquetas tanto en el momento de creación del DLO, como *a posteriori*.
- c) Finalmente, porque el criterio de la forma de trabajo (servidor o cliente), además del interés para los creadores de metainformación a la hora de seleccionar una herramienta que les ayude a la asignación de metadatos, es un criterio suficientemente amplio como para analizar, dentro de esa división, toda una casuística de herramientas que exponemos a continuación.

4.3.2.1 Plantillas

Las plantillas para la creación de metadatos son mucho más sofisticadas y completas que las destinadas a la simple creación de etiquetas <META>. Son algo más que meros formularios Web, normalmente, CGIs basados en Web, *applets* de Java o ventanas de Visual Basic o de Java, que se abren independientemente del navegador, y que funcionan en modo servidor. En la mayoría de los casos, son producto de proyectos de investigación en metadatos realizados por las principales instituciones que participan en el desarrollo de los grandes modelos de metainformación y que creen en la aplicación de esta tecnología para el diseño de sus

sistemas de información electrónica. Analizamos a continuación los casos más importantes de este tipo de herramientas.

The Nordic Metadata Project Template es una de las plantillas más antiguas, creada en septiembre de 1997 por Traugott Koch y Mattias Borell, en el seno del *Proyecto Nórdico de Metadatos*, uno de los primeros intentos de aplicación y desarrollo del modelo Dublin Core. Se trata de una sencilla herramienta, un formulario donde se puede completar toda la información descriptiva relativa a un DLO o recurso Web, siguiendo los quince elementos que conforman el DC. Dentro de las características principales de esta plantilla, podemos destacar que:

- Proporciona una ayuda en línea detallada, tanto del funcionamiento de la plantilla como del propio modelo DC (cada uno de los elementos del formulario tiene un enlace a una explicación con ejemplos de su uso en el DC).
- Dispone de una versión simplificada de la plantilla <<http://www.lub.lu.se/cgi-bin/nmdc.pl?simple=1>> a un nivel mínimo de metadatos recomendados, para aquellos casos en los que no es necesario un nivel de detalle tan preciso. No obstante, el usuario puede decidir tanto repetir un elemento, como no incluirlo en su descripción.
- Admite varios calificadores del DC. Por ejemplo, a la hora de la creación de metadatos relativos a las materias (DC.Subject), además de la introducción de palabras clave, también permite caracterizar el contenido de los documentos descritos a través de distintos *schemes*, vocabularios controlados o como clasificaciones, esquemas elegidos por el usuario de entre dieciséis vocabularios controlados (p. ej. LCSH, u otras clasificaciones más específicas de los países participantes en el *Nordic Metadata Project*, como la *Finnishe Forest SH*, etc.) y diecisiete clasificaciones (p. ej. DDC, LCC, o la SAB Class de Suecia, o la Clasificación Decimal Danesa DK5). De esta forma, se pueden describir las materias de forma controlada gracias al atributo SCHEME, como muestra el ejemplo a continuación:

```
<META NAME="DC.subject" CONTENT="Dublin Core, metadata,
resource discovery">
<META NAME="DC.subject" SCHEME="LCSH" CONTENT="Cataloging of
computer files">
<META NAME="DC.subject" SCHEME="UDC" CONTENT="518.118">
```

Esta posibilidad de calificar, a través de un esquema (*scheme*), la descripción por materias de los DLOs, fomentará la constitución de sistemas de recuperación basados en metadatos, más precisos.

- Otra particularidad destacable es que dispone de un enlace interactivo a un servicio de URNs (*Uniform Resource Names*) para los DLOs finlandeses, noruegos o suecos¹³⁴, como un valor añadido a la plantilla.
- Una vez completada toda la información, genera los resultados bien para previsualización y comprobación de cada una de las etiquetas (Fig. 24), bien para la inclusión del código fuente en un documento HTML 3.2. o 4.0.

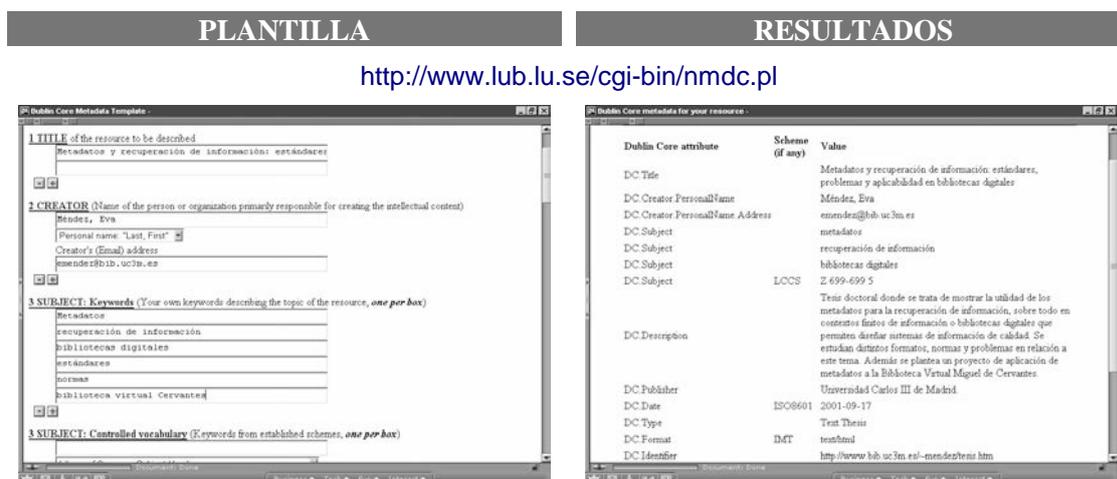


Fig. 23. Nordic Metadata Project Template

Fig. 24. Nordic Metadata Project (resultados)

¹³⁴ El servicio al que hacemos referencia es el *Nordic Countries URN generator* <http://www.lub.lu.se/cgi-bin/nmurn.pl?mode=dccreator&lang=en&form=dcform&field=F_15_1> también se desarrolló en el marco del *Nordic Metadata Project*, como complemento a su plantilla de metadatos y a fin de normalizar las descripciones de DLOs noruegos, suecos y finlandeses. La sintaxis de los URNs producidos está autorizada por el Grupo de trabajo sobre el URN del IETF.

En el ejemplo se puede ver la introducción de la metainformación descriptiva de esta tesis doctoral como si fuese un documento publicado en Internet, según una serie de campos repetibles que se corresponden con las etiquetas que conforman el esquema del Dublin Core (Fig. 23); asimismo se ve la presentación en pantalla de la previsualización de la metainformación creada.

```
<META NAME="DC.Title" CONTENT="Metadatos y recuperaci&oacute;n  
de informaci&oacute;n: est&aacute;ndares, problemas y  
aplicabilidad en bibliotecas digitales">  
<LINK REL=SCHEMA.dc  
HREF="http://purl.org/metadata/dublin_core_elements#title">
```

A pesar de la sencillez y de la adecuación de esta plantilla, tiene algunas limitaciones, de las cuales la más evidente es que sólo sirve para generar metadatos Dublin Core. Otras limitaciones más concretas son, por ejemplo, que los calificadores de lengua están limitados a la norma ISO639-1, que no existen calificadores de lengua para cada uno de los elementos, que tiene un sesgo marcado hacia la descripción de recursos electrónicos nórdicos, o que sólo permite HTML como sintaxis de codificación. Esta plantilla es una buena opción para crear metadatos según el modelo DC, y por ello ha servido de modelo para inspirar otras. Por ejemplo, la versión sueca de esta plantilla <http://safari.hsv.se/metadata/safari.pl?lang=sv>, adaptada por la base de datos de recursos de investigación, SAFARI, o la plantilla de la Biblioteca Nacional de Dinamarca <http://www.pligtaflevering.dk/anmeld/anmeld.pl> o la versión alemana adaptada por la BSZ <http://www.bsz-bw.de/diglib/medserv/konvent/metadat/dcmake1t.html>, que incluso le han dado un nombre propio *DC-Meta-Maker des BSZ*.

Si en primer lugar hemos analizado la más antigua dentro de los modelos de plantilla, ahora vamos a hablar de la herramienta más completa, desde nuestro punto de vista, para la creación de metadatos a través de una aplicación que funciona en línea: **Reggie**. Se trata de un sencillo editor desarrollado por el Centro de Tecnología de Sistemas Distribuidos (DSTC) que facilita la creación de metadatos basados en distintos formatos, a través de una aplicación Java en varios idiomas. Permite completar los campos, a través de un *applet*, como si fuese una base de datos tradicional y posteriormente se pueden visualizar el código resultante, bien a través

del navegador o bien enviándolo por correo electrónico, tanto en formato HTML 3.2 ó 4.0, como en RDF. Permite la creación de metadatos según el formato DC, en varios idiomas (inglés, italiano, finlandés, francés, alemán, noruego, español y portugués). Además crea un *applet* de las mismas características, pero sólo inglés, para la creación de otros modelos, como p. ej., GILS, IMS, GEM para los metadatos de la *National Library of Education* de Australia (NLE). Además de estos esquemas, Reggie, facilita la creación de otros modelos *ad hoc*, bien basados en formatos predefinidos como el Dublin Core, o estrictamente nuevos¹³⁵.

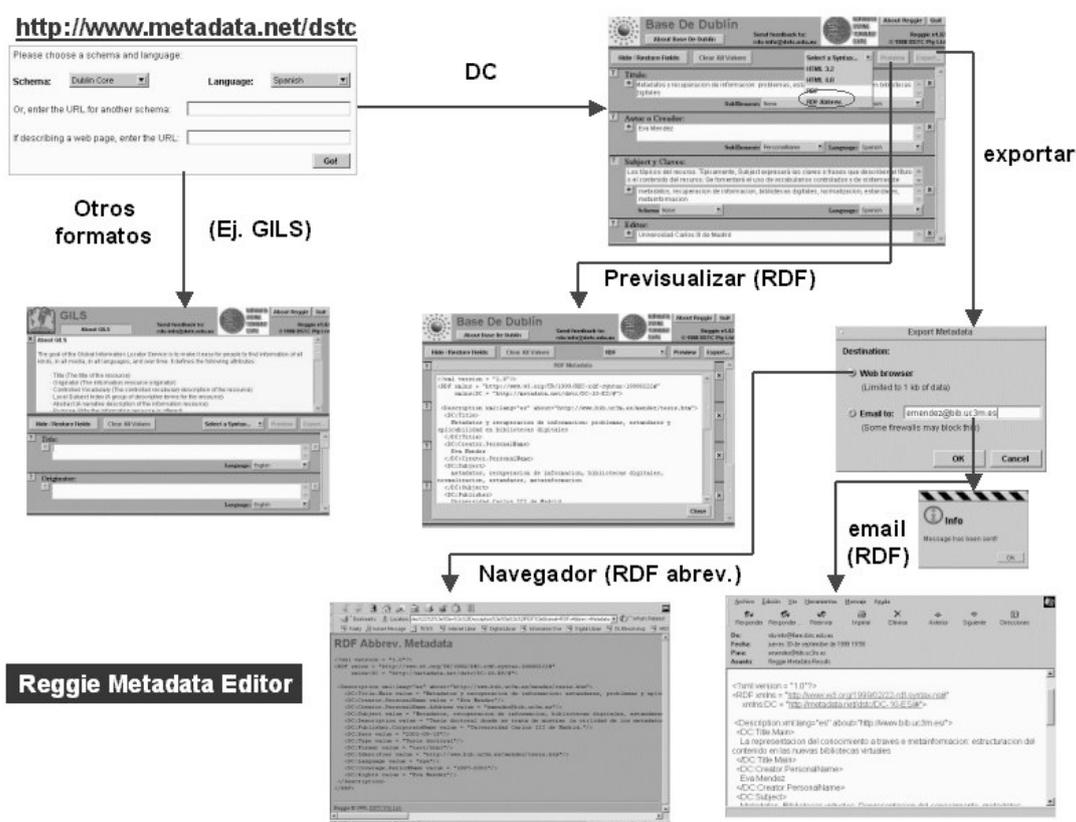


Fig. 25. Proceso de trabajo con Reggie Metadata Editor¹³⁶

¹³⁵ Para más información sobre la creación de esquemas (*schemas*) nuevos, distintos de los que contempla la aplicación por defecto (Dublin Core, GILS, ANZLIC, AGLS, EdNA, IMS, GEM y vCard), Vid. <http://metadata.net/dstc/SchemaFiles.html>

¹³⁶ Sobre el funcionamiento de este editor de metadatos es interesante consultar el resumen práctico planteado por Brian Kelly y Andy Powell para el taller titulado: *Institutional Web Management* [cont.]

Para evaluar esta herramienta, nuevamente recurrimos al supuesto que esta tesis fuese un DLO para la WWW (la Fig. 25 muestra todo el proceso de trabajo con esta aplicación). En primer lugar, se selecciona el esquema y el idioma de trabajo (en el este caso, Dublin Core y español), a pesar de que, como hemos explicado, este programa permite trabajar con otros formatos sólo en inglés (en la figura se recoge el caso de la interfaz de trabajo para el formato GILS); una vez completados todos los campos, se puede previsualizar el código resultante con la información que hemos introducido, así como la exportación del código en RDF¹³⁷ abreviado (también podría haberse seleccionado en RDF normal, HTML 3.2 o HTML 4.0); finalmente la visualización del código generado por Reggie se presenta bien en el Navegador, bien por correo electrónico, pudiéndose, en cualquiera de los casos, copiar y pegar en el documento Web concreto para el que hemos asignado los metadatos. Durante todo el proceso de trabajo con Reggie tiene una ayuda contextual tanto sobre el esquema de metadatos con el que se está trabajando, como sobre el valor de cada uno de los elementos.

A pesar de que consideramos que Reggie es la plantilla más completa para la edición o creación de metadatos, también presenta algunas restricciones como: la necesidad de un navegador de última generación (Netscape 4.x, o superior e Internet Explorer 4.x o superior, más rápido con Netscape), no admite vocales acentuadas a pesar de que la interfaz elegida sea, en el caso del modelo DC, el español, o el hecho de no poder copiar datos de la ventana del *applet* a otras ventanas u otras

Workshop que tuvo lugar en la Universidad de Newcastle el 16 de Septiembre de 1998: <http://www.ukoln.ac.uk/web-focus/metadata/seminar-materials/exercises/reggie/reggie.pdf>. Sin embargo es preciso comentar que, aunque el funcionamiento de Reggie no ha variado substancialmente, este material que citamos esta basado en la versión anterior de esta herramienta (Septiembre 1998).

¹³⁷ La importancia de esta aplicación Java para la creación de metadatos se puede ver en la asunción de la sintaxis RDF, incluso en 1998 antes de que fuese una recomendación del W3C (*Cfr.* 5.2.3). Esto hace que algunos de los autores consagrados ya por sus trabajos sobre XML, utilicen Reggie para hablar de la utilización de RDF. *V. gr.* Simon St. Laurent, Robert Biggar. *Inside XML DTDs*. New York, etc.: McGraw Hill, 1999, p. 221 y ss. (Figures: 12-9, 12-10, 12-11 y 12-12).

aplicaciones en curso¹³⁸, sin embargo el envío de los metadatos a través del correo electrónico, o su presentación en el navegador, permite copiar el código resultante, en la cabecera de cualquier documento HTML o XML, ya que una de las características más destacables de esta aplicación es que soporta ambas sintaxis.

Otra herramienta interesante, también creada por el DSTC es **Reg**, definido como el editor "ligero" de metadatos (*Lightweight metadata editor*), por contraposición a Reggie, que es "el editor de metadatos" (*The metadata editor*). También se trata de una aplicación, que como la anterior, genera metainformación según distintos modelos o esquemas (concretamente, DC, en sus tres variantes, AGLS, EdNa y QLDGOV¹³⁹) (Fig. 26) y distintas sintaxis (RDF —sintaxis completa o abreviada—, HTML —versiones 3.x y 4.x— y SOIF). La plantilla de Reg está basada en un lenguaje de *scripts* propio <<http://flare.dstc.edu.au/reg/reg.js>>, y tiene una apariencia y funcionamiento similar a la plantilla del *Nordic Metadata Project*, que analizamos en primer lugar en esta sección. Como en ella, Reg tiene un fichero de ayuda en línea <<http://flare.dstc.edu.au/reg/help/help-dc.html>>, que remite a la explicación del uso de cada uno de los elementos, también se pueden repetir campos del formulario (que no son otra cosa que etiquetas <META>), o eliminarlos, y al final de la plantilla permite exportar o salvar los metadatos según diferentes sintaxis de salida.

¹³⁸ Esto se debe principalmente a la seguridad Java que tienen la mayoría de los navegadores. En la versión de Reggie que hemos analizado (ver. 1.68, diciembre 1998, la última en el momento de redacción de esta tesis doctoral) no se pueden copiar y pegar los datos entre otras aplicaciones Windows y el *applet* de Reggie, sin embargo se anuncia la posibilidad de corregir esta limitación en próximas versiones. En nuestra opinión, Reggie quizás no se ha desarrollado como esperábamos debido a que el DSTC se ha comprometido posteriormente en proyectos más sólidos y complejos como Harmony <<http://metadata.net/harmony>>, un proyecto internacional iniciado en Julio de 1999 para analizar cuestiones relacionadas con la descripción de toda la complejidad multimedia de los DLOs en las nuevas bibliotecas digitales.

¹³⁹ Es decir, el modelo DC y en algunas de las variantes a este formato, surgidas en distintos ámbitos de información en Australia: AGLS (*Australian Government Locator Service*), EdNA (*Educational Network Australia*), QLDGOV (*Queensland Government Metadata Standard*).

Reg utiliza distintos esquemas de metadatos de tal forma que la aplicación puede adaptarse a distintas necesidades del usuario. Hemos mencionado los modelos por defecto que contempla, pero además, se pueden proponer esquemas *ad hoc*. Cada fichero de esquema consiste en una serie de etiquetas `Description` dentro de una etiqueta `RDF`; cada etiqueta `Description` tiene un atributo `ID` que se usa para identificar el elemento de metadatos; los distintos elementos se mostrarán en el orden que aparecen en el esquema. Un ejemplo sencillo de los esquemas que se manejan en Reg es¹⁴⁰:

```
<RDF>
<Description ID="Subject">
<REG:prefix>DC</REG:prefix>
</Description> </RDF>
```

<http://flare.dstc.edu.au/cgi-bin/reg/demo.cgi>



Fig. 26. Reg (opciones)

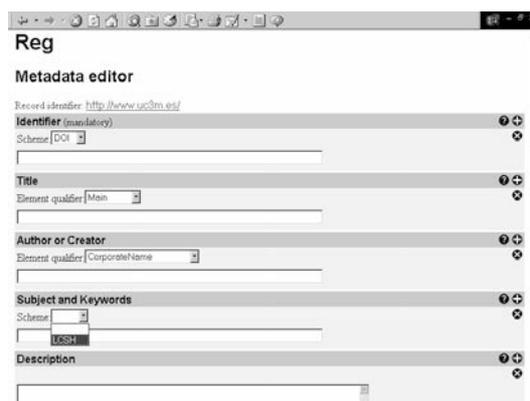


Fig. 27. Reg (plantilla según DC)

Además de la flexibilidad que proporciona Reg al trabajar con esquemas RDF/XML es capaz también de trabajar con metadatos embebidos en recursos HTML y utilizarlos como valores iniciales del registro de metadatos que se creará y completará a partir de ellos. Para trabajar de esta forma y aprovechar los metadatos

¹⁴⁰ El ejemplo muestra el esquema de la etiqueta `DC.Subject` del Dublin Core. Para ver el esquema completo del modelo DC en Reg, Vid. <http://flare.dstc.edu.au/reg/schemas/schema-dc.xml>

incluidos en una página, se debe introducir el URL como identificador¹⁴¹. A pesar de todo esto, Reg también tiene limitaciones funcionales, como por ejemplo el navegador, sólo funciona correctamente en Internet Explorer 5.x, no puede visualizarse en Netscape 6.x, y en Opera 5.x aparece la plantilla incompleta.

Al igual que Reggie, Reg es una plantilla utilizada en servicios de información australianos para crear sus metadatos, como por ejemplo en AVEL (*Australian Virtual Engineering Library*) <<http://avel.edu.au>>. AVEL tiene su propia adaptación de la plantilla Reg <<http://avel.library.uq.edu.au/cgi-bin/reg/reg.cgi?>> de tal forma que la plantilla se ajusta al propio esquema de AVEL, con sus campos propios (AVEL.Publisher, AVEL.Collection, AVEL.ResourceType), sus variantes del DC (Coverage.PeriodName, Coverage.PlaceName) o sus particularidades de calificadores (como por ejemplo, en la etiqueta Subject, se selecciona entre el esquema LCSH, también en Reg, y su propio vocabulario para el control de materias AVEL). Otras características de esta adaptación de Reg para crear metadatos en una biblioteca digital concreta como AVEL son, por ejemplo, que sólo permite salvar los registros, previa introducción del *login* y *password*, y no exportarlos en diferentes sintaxis como veíamos en la versión original de la plantilla Reg, y que el fichero de ayuda contextual en línea también es sólo accesible para los usuarios registrados de AVEL. Con todo, esta adaptación de la plantilla de Reg a una biblioteca digital (*subject gateway*) concreta, demuestra su flexibilidad y su utilidad para la creación y gestión de metadatos.

A pesar de lo interesante de Reggie y Reg, la plantilla para la creación de metadatos por excelencia, al menos la más conocida y divulgada en el ámbito del Dublin Core, es **DC-Dot**. Desarrollado por la UKOLN, DC-Dot es un servicio Web

¹⁴¹ En el ejemplo de la Fig. 26, se ha introducido el URL de la *homepage* de la Universidad Carlos III <<http://www.uc3m.es>>, pero el sistema no ha extraído ninguna metainformación a pesar de que dicha página contiene las etiquetas HTML: <TITLE>, <META NAME="description"> y <META NAME="keywords">.

editor de metadatos que genera etiquetas Dublin Core y las visualiza tanto en sintaxis HTML como RDF/XML (desde junio del 2000, también en XHTML). Así pues, según la clasificación del DTSC que apuntábamos al principio de este apartado, DC-Dot es al mismo tiempo un editor y un generador de metadatos.

La forma de trabajar de este programa es muy sencilla: el usuario introduce en un formulario el URL del DLO al que quiere asignar metadatos, ya sea este un documento HTML, un fichero de Word o una presentación de PowerPoint¹⁴². En primer lugar extrae automáticamente toda la metainformación posible de la página, esto es, al menos la información relativa a: la fecha¹⁴³ (DC.Date), formato del documento (DC.format), URL (DC.Identifier), y si hemos seleccionado la opción de que determine automáticamente al editor, también la etiqueta DC.Publisher. Esta información mínima es la que extrae por ejemplo, cuando el URL que pretendemos analizar es de una imagen (Fig. 28), en la mayoría de los casos, es decir, páginas Web, tratará de extraer otra información basándose para ello

¹⁴² A pesar de que en la presentación de esta herramienta <<http://www.ukoln.ac.uk/metadata/dcdot>> sólo se habla expresamente de la generación de metadatos DC para documentos ofimáticos convencionales como Word y PowerPoint, también hemos podido comprobar, que crea metadatos (básicos y obviamente no extraídos en ningún caso del código fuente) también para otros formatos de ficheros que pueden tener los distintos DLOs de la Web, como por ejemplo, .pdf, o formatos de imágenes, como .jpg o .gif. La Fig. 28 presenta los metadatos extraídos para la imagen <http://www.bib.uc3m.es/~mendez/fotos/yo.jpg>

¹⁴³ La precisión de este software-plantilla se ve por ejemplo en la extracción de la fecha, ya que independientemente de la fecha que aparezca en el contenido del documento, si no existe este elemento de metadatos, extrae la de la última transferencia al servidor entendiéndolo por ello —como hacemos nosotros al citar los recursos electrónicos de esta tesis— que esa es la última fecha de edición/reedición. Por ejemplo, si utilizamos DC-Dot para crear los metadatos del recurso (DLO): <http://web.simmons.edu/~schwartz/meta.html> extrae como fecha 2001-06-29, mientras que en el texto del documento aparece la expresión *Last revised June 16, 2001* y los metadatos embebidos en la cabecera del mismo, no aparece la etiqueta relativa a la fecha. Esta característica de DC-Dot nos ha servido además en la metodología de este trabajo para consignar las fechas de revisión de los documentos electrónicos que citamos, sobre todo en el caso de que no apareciera ninguna. Cfr. Capítulo 2, nota 17-d.

en la etiqueta <TITLE> o cualquier metaetiqueta (*keywords*, *description*, etc.) que contenga la cabecera del documento.

En la misma pantalla aparece una plantilla, en este caso con todos los elementos del Dublin Core, para completar la metainformación relativa al DLO, donde también se puede elegir el formato de visualización de los metadatos creados (HTML, por defecto, RDF o XHTML) y el lenguaje de las etiquetas (inglés o francés). Además, una vez generados los metadatos, según uno de los tres esquemas de visualización que hemos señalado, DC-Dot actúa además como conversor de metadatos DC a otros formatos (USMARC, SOIF, IAFA/ROADS, TEI Header, GILS, IMS, XML e incluso OLSTF —formato propietario del proyecto NewsAgent¹⁴⁴—), permite guardar el archivo resultante en un directorio temporal del ordenador, en el lenguaje de codificación que requiera cada formato.

Además de su presencia en la Red como formulario generador de metadatos en el URL: <http://www.ukoln.ac.uk/metadata/dcdot>, DC-Dot se admite la descarga en forma de Script-CGI, que fue la primera versión en Perl del programa (julio de 2001); los ficheros de la aplicación Java (Java DC-Dot) también se pueden descargar¹⁴⁵. Permite arrastrar un marcador¹⁴⁶ a la barra personal del navegador, de tal forma que

¹⁴⁴ NewsAgent es un proyecto sobre revistas electrónicas dentro del programa eLib, que ya hemos mencionado. Este proyecto trabaja con Dublin Core y con el formato propio OLSTF. Un registro OLSTF consiste en un conjunto de valores distribuidos en etiquetas (una por línea) de dos letras, al modo de las bases de datos bibliográficas (TI —título—, AB —resumen—, etc.) seguidas del valor correspondiente. Este proyecto, tiene además su propia versión del editor DC-Dot: NewsAgent Metadata Creator <<http://www.ukoln.ac.uk/metadata/newsagent/dc>> que genera una plantilla específica donde se señalan los campos obligatorios en NewsAgent para convertirlos a registros OLSTF. Más información sobre este proyecto en: <http://www.sbu.ac.uk/litc/newsagent>

¹⁴⁵ Ambas versiones del software, se pueden descargar comprimidas en: <ftp://ftp.ukoln.ac.uk/metadata/tools/dcdot>

¹⁴⁶ Lo que colocamos en la barra personal de nuestro navegador, es concretamente este *script* de Java:
,
[cont.]

siempre está activada la posibilidad de analizar, generar o editar los metadatos de cualquier página que estemos visualizando.

En cuanto a la ayuda, a pesar de que DC-Dot presenta varios enlaces que explican el contenido de las etiquetas, la utilización del programa, etc. es preciso destacar otra herramienta desarrollada por Andy Powell, también en el seno de la UKOLN, denominada **DC-assist** y que tiene como finalidad exactamente eso: asistir en la creación de metadatos Dublin Core¹⁴⁷ (Fig. 29).

<http://www.ukoln.ac.uk/metadata/dcdot>

<http://www.ukoln.ac.uk/metadata/dcassist>

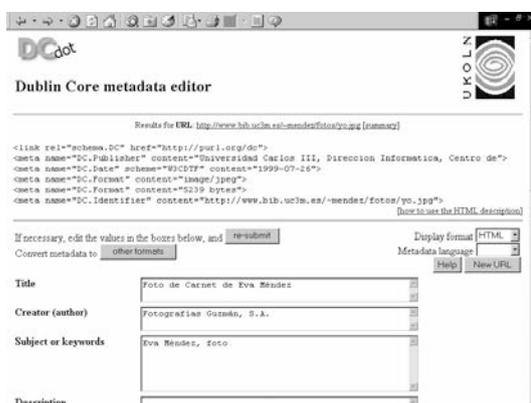


Fig. 28. DC-Dot. Metadatos extraídos automáticamente y plantilla de edición¹⁴⁸



Fig. 29. DC-assist v.1.6. (Elemento coverage)

que permitirá extraer y editar los metadatos de cualquier página que estemos visualizando en él. Esta forma de trabajar que tienen determinadas herramientas basadas en la Web es bastante frecuente, por ejemplo Klarify. Esto da la posibilidad y utilidad de tener potencialmente operativa la aplicación durante la navegación.

¹⁴⁷ No sólo a los creados con DC-Dot, sino en general, a todas las aplicaciones que contemplan este esquema de metadatos. La propia DCMI (*Dublin Core Metadata Initiative*) <<http://www.dublincore.org/tools>> clasifica a DC-assist, como una "utilidad", dentro de la categorización que hace de las distintas herramientas que trabajan con metadatos.

¹⁴⁸ Sobre el funcionamiento de este editor también existe un documento práctico-resumen de realizado también por Brian Kelly y Andy Powell en el taller que mencionamos, *Institutional Web Management Workshop*: <http://www.ukoln.ac.uk/web-focus/metadata/seminar-materials/exercises/dc-dot/dc-dot.pdf>

DC-assist es una aplicación JavaScript que presenta un listado con las quince etiquetas del DC a las cuales acompaña información sobre sus calificadores, una concisa definición, un comentario relativo a su uso, el *namespace*¹⁴⁹ correspondiente, ejemplos, enlaces a otra información (como por ejemplo, los esquemas de codificación que pueden utilizarse para una etiqueta en particular¹⁵⁰) y se pretenden incluir también, guías para la entrada de datos. A pesar de que, como la mayoría de las herramientas relacionadas con la metainformación, surge en el seno del modelo DC y concretamente como ayuda a la plantilla del DC-Dot, DC-assist es muy flexible y se puede adaptar para utilizarse con cualquier otra aplicación o para su uso en local a través de la creación de un fichero de configuración <<http://www.ukoln.ac.uk/metadata/dcassist/configuration>>; asimismo, en un futuro, toda la información posible se podrá extraer de un perfil de aplicación (*application profile*) almacenado en un registro de metadatos¹⁵¹, directamente de un esquema RDF o de una combinación de ambos.

Existen otras muchas plantillas como las que hemos descrito en las páginas anteriores, todas ellas sencillos formularios para completar el contenido de los metadatos. En general, son semejantes a las que hemos analizado, pero normalmente asociadas a un proyecto de organización de un tipo de información o sobre un tema concreto, como por ejemplo: o *Medical Metadata Creator* <<http://medir.ohsu.edu/bicc-informatics/ebm/latest.htm>> para la creación de metadatos de recursos de información médica o *My MetaMaker 1.31* <<http://physnet.uni-oldenburg.de/services/mmm>> para

¹⁴⁹ Vid. 5.2.3. y Glosario.

¹⁵⁰ V. gr. en la Fig. 29 aparecen enlaces que explican los distintos esquemas que pueden afectar a la etiqueta DC.Coverage, como por ejemplo el formato de codificación de fecha del Consorcio Web (W3CDTF), ISO 8601:1988.

¹⁵¹ Por ejemplo el registro desarrollado por el proyecto SCHEMAS <<http://www.schemas-forum.org/registry/registry.html>>. Todas estas cuestiones de índole técnico, necesarias para explicar el funcionamiento de estas aplicaciones, se matizarán suficientemente a lo largo del próximo capítulo. Cfr. 5.2.1.2. y Glosario, *application profile*, *metadata registry*.

DLOs sobre física, ambas plantillas basadas en formato del Dublin Core. Existen también visiones nacionales como la plantilla danesa *Metadata Creator Version 2.0–Publikationer* <http://www.nnc.dk/meta_pub.htm>. También es frecuente que se creen plantillas para la asignación de metadatos según un formato o esquema particular. Este es el caso por ejemplo, de **RSS-Express** <<http://rssxpress.ukoln.ac.uk>>, creada también por la Oficina para el Funcionamiento en Red de Bibliotecas e Información del Reino Unido (UKOLN) para la creación, modificación y registro de canales de noticias RSS¹⁵².



Fig. 30. RSS-Express. Canal Xml.com

La UKOLN tiene registrados, actualmente (julio 2001) en su directorio, veintinueve canales de noticias, tales como Freshmet, UKOLN News, Xml.com, Slashdot, etc. que pueden editarse (Fig. 30) o visualizarse. Al ser RSS una aplicación

¹⁵² RSS (*RDF Site Summary*) es un formato de metadatos originalmente desarrollado por Netscape para el desarrollo de contenidos en portales de información Web. Desde el 6 de diciembre del 2000, es una especificación estable (1.0) para la "sindicación" de contenidos, que utiliza la sintaxis RDF y el esquema del DC para describir el contenido de las noticias agregando (*syndicate*) el resumen del sitio a la información de titulares en el Web. Vid. Rael Dornfest. Writing RSS 1.0 [documento HTML]. *O'Reilly Network*, 25 de agosto de 2000. Disponible en: http://www.oreillynet.com/pub/a/network/2000/08/25/magazine/rss_tut.html (consultado el 28 de agosto de 2000). Vid. Glosario, *syndication/syndication format*.

XML adaptada a la especificación RDF del W3C, RSS-Express permite generar el código RDF que facilitará que los metadatos asociados a titulares se puedan intercambiar entre los proveedores de contenidos de noticias y los portales. El resultado de la edición con RSS-Express es un documento .rdf con la siguiente estructura:

```

XML Declaration
Container(Declaración de namespaces):
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:syn="http://purl.org/rss/1.0/modules/syndication/"
  xmlns:taxo="http://purl.org/rss/1.0/modules/taxonomy/">
  Channel Description (descripción del contenido del
  canal, incluyendo el listado de ítems embebidos en
  <rdf:Seq> y de las imágenes, en su caso)
  Image Description (opcional)
  Item Description
  ...
  Item Description (hasta 15 ítems)
  Text Input Description (opcional)
  
```

Otra plantilla de este mismo tipo, es decir, una aplicación que facilita la creación en línea de metadatos según un formato específico, en este caso (*AGLS Australian Government Locator Service*) es **BetaMeta Bureau** <<http://www.bettameta.com>>.

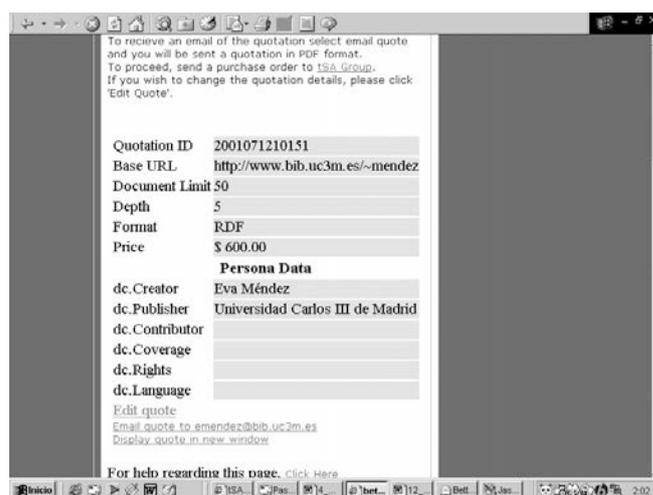


Fig. 31. BetaMeta Bureau

BetaMeta tiene además otras características destacables, en relación con el resto de plantillas analizadas: se trata de una herramienta de pago¹⁵³ a pesar de que funciona en modo servidor, y la creación de metainformación es automática pues pertenece a la empresa australiana tSA que ha desarrollado **Klarity** (*Vid. supr.*) y por ello utiliza ese software para la indización automática de DLOs y para la representación de sus metadatos. Lo que permite esta plantilla es generar un presupuesto¹⁵⁴ (Fig. 31) para la creación de metadatos de un sitio Web determinado, introduciendo para ello una información mínima: el URL, el nivel de profundidad para seguir enlaces a partir de dicho URL, el número de páginas que conforman el sitio, el formato de salida de la metainformación (HTML, RDF o BEPXML), y unos datos personales mínimos según el formato DC (creador, editor, contribuidores, cobertura, derechos e idioma, siendo obligatorios sólo los dos primeros). Una vez generado el presupuesto se crea un identificador y se envía por correo al usuario en formato .pdf, y en el caso de que este lo acepte (en el plazo máximo de siete días), se procesarán las páginas Web y se generará un fichero de salida en formato .zip que podrá descargar el usuario.

Con todo, y a pesar de la variedad y especificidad creciente de estas plantillas o servicios de edición de metadatos, podemos decir que, en general, son herramientas sencillas, dirigidas a un creador de metadatos que no tiene que ser especialista ya que son aplicaciones servidor muy guiadas como se ha visto en los ejemplos elegidos. Las aplicaciones informáticas que realmente permitirán realizar una gestión completa

¹⁵³ Para evaluar esta herramienta hemos tenido que registrarnos para obtener una password. www-data@topic.com.au. *Bettameta Bureau Site Password* [correo electrónico personal], 16 de mayo de 2001; 20:58. Disponible en: emendez@bib.uc3m.es

¹⁵⁴ Por ejemplo, la creación de metadatos para el sitio Web del Dpto. de Biblioteconomía y Documentación de la Universidad Carlos III <<http://www.bib.uc3m.es>> (Quote No. 2001071210151), seleccionando un nivel de profundidad de seguimiento de enlaces 5, y con un máximo de 50 páginas Web, tiene un precio de \$600 (120.000 pts. aprox.), que podemos tomar como indicador del coste de la creación de metadatos y por ende, como justificación de que la mayor parte de los sitios no utilicen esta información sobre la información.

de sistemas de metainformación complejos y crear, por ello, verdaderas bibliotecas digitales son, generalmente software comerciales que trabajan en los ordenadores personales o en los servidores de una organización cuya información se distribuye en modo Web (Internet o Intranet).

4.3.2.2. Software cliente

Las distintas aplicaciones que funcionan en modo cliente son, a diferencia de las plantillas para la creación de metadatos (4.3.2.1.) y del software cliente para la asignación de metaetiquetas (4.3.1.2), en la mayor parte de los casos, herramientas o módulos de software comerciales. Por esta razón, lo que hemos podido evaluar son demos de los programas o a lo sumo la información cedida por los desarrolladores de los mismos¹⁵⁵.

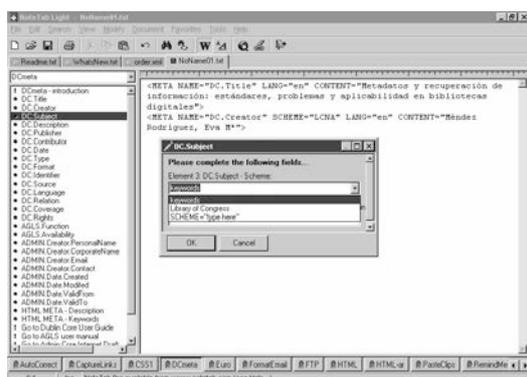


Fig. 32. Creación de metadatos con NoteTab

Antes de analizar el software comercial debemos estudiar qué opciones se presentan para trabajar con metadatos en forma de software cliente. La primera de ellas es la de trabajar con una herramienta genérica de edición de páginas Web adaptada a las necesidades específicas de metadatos. En este sentido es paradigmática la adaptación hecha por

¹⁵⁵ Como anticipamos al hablar de la metodología de este trabajo (Cfr. Capítulo 2, nota 11), en el caso de MetaStar, que no dispone de Demo comercial, hemos accedido a la información de la base de conocimiento (*Knowledge Base*) de Blue Angel Technologies a través de un *login* y *password* cedidos por la propia empresa, para poder evaluar dicha aplicación. Chris Hynes. Re: METASTAR [correo electrónico personal], 14 de mayo de 2001; 14:34. Disponible en: emendez@bib.uc3m.es

la *State Library of Tasmania* de la aplicación genérica **NoteTab**¹⁵⁶ al que se le añade un fichero denominado *Dublin Core Clipbook for Super NoteTab* (DCmeta.clb¹⁵⁷) para poder asignar fácilmente metadatos según los estándares DC, AGLS y A-Core (esquema australiano basado en gran medida en el DC que incluye una serie de elementos accesorios para controlar la circulación, autoría e integridad de los metadatos). Al añadir dicho archivo, NoteTab, que en principio no es más que un completísimo editor textual de páginas Web creado por la empresa suiza Fookes Software, agrega a sus funciones un listado de etiquetas, que una vez seleccionadas, proporciona un cuadro de diálogo a través del cual se puede completar la información de cada elemento de metadatos, incluyendo el *scheme*¹⁵⁸ y el contenido. El hecho de incluir esta función específica, no obsta para que, al igual que otros editores de páginas Web, NoteTab permita crear las etiquetas <META> convencionales del HTML, así como las propiedades del documento. La ventaja de utilizar un editor genérico de páginas Web al que se le ha añadido, en este caso, toda la información sobre los distintos modelos, es que crea las cabeceras como una parte más del documento Web, aunque los metadatos también se pueden almacenar en un fichero aparte.

¹⁵⁶ Programa-versión: NoteTab™ Light 4.85
Homepage: <http://www.notetab.ch>
Tipo: Freeware
Plataforma: Win 95/98/NT/2000/ME
Nombre (tamaño) del fichero: ntfree.zip (1506 Kb)
Descarga: <http://www.notetab.ch/ftp/ntfree.zip>
Fecha de descarga: 30/06/2001 20:48

¹⁵⁷ Este fichero (*Dublin Core Clipbook for Super NoteTab v4.00*), creado por Lloyd Sokvitne, un bibliotecario del Servicio *Tasmania Online* <<http://www.tas.gov.au>> para asistir en la creación y adición de metadatos, se puede obtener en: <http://www.dstc.edu.au/Research/Projects/metaweb/download/DCmeta.clb> Sobre cómo incluir el fichero en la aplicación NoteTab Vid. <http://www.dstc.edu.au/Research/Projects/metaweb/download/DCinstall.txt>

¹⁵⁸ Vid. Glosario.

Otra solución son las aplicaciones creadas completamente *ex novo* para asignar y gestionar metadatos. En este sentido, en el ámbito español destaca una herramienta creada por el grupo de trabajo de Red Iris IRIS-INDEX, denominada **MetaWebber**¹⁵⁹, que toma una página HTML generada con cualquier editor y permite incorporarle etiquetas HTML con metainformación corporativa. Para ello se basa en un fichero de plantilla donde se almacenan los metadatos. Esos ficheros con la misma metainformación, se guardan en cada directorio y afectan a todos los ficheros que pertenezcan al mismo directorio. Este sistema, pertenece a un proyecto denominado Webber¹⁶⁰ dentro del Plan Nacional de Aplicaciones y Servicios Telemáticos, titulado *Sistemas de integración y extracción de metainformación para servidores de red*. Webber es, según sus creadores¹⁶¹, *un entorno basado en componentes para armonizar el contenido de servidores web e incluir automáticamente metainformación en él*. Originalmente concebido como un preprocesador de HTML, ha evolucionado hasta llegar a ser un sistema completo de producción y mantenimiento de contenido y metainformación para páginas Web. El

¹⁵⁹ MetaWebber, versión 1.0.6 beta (17/01/2000): <http://www.rediris.es/si/iris-index/herramientas/metawebber>

¹⁶⁰ En realidad MetaWebber es un módulo que, junto con Webber 1.0 y Webbo conforman el *freeware* Webber, cuya última versión en el momento de depósito de esta tesis doctoral es la 1.1.0. build 18 del 02/05/2001, y se puede descargar del servidor FTP de RedIris: <ftp://ftp.rediris.es/rediris/webber/webber-1.1.0-b18.tar.gz>

¹⁶¹ Webber [documento HTML]. RedIris, rev. 2 de mayo de 2001. Disponible en: <http://www.rediris.es/app/webber> (consultado el 15 de junio de 2001). Para más información sobre este sistema, *Vid.* Diego R. López. Webber: Uso de componentes para la armonización de contenidos y metadatos. En: *ISKO (5. 2001. Alcalá de Henares). La representación y organización del conocimiento: Metodologías, modelos y aplicaciones* [CD-ROM]. Alcalá de Henares: Universidad de Alcalá de Henares, 2001, p. 73-81 y también Diego R. López y Javier Masa, *Op. cit.*, <http://www.rediris.es/rediris/boletin/45/enfoque1.html>

código fuente de la cabecera de metainformación generado con este sistema, es siempre muy uniforme y tiene el aspecto que presenta este ejemplo¹⁶²:

```
<!doctype html public "-//W30//DTD W3 HTML 3.0//ES">
<!-- -----
Página generada automáticamente con Webber 1.1.0
N O E D I T A R
© 1996-2000 Centro de comunicaciones CSIC/RedIRIS
----- -->
<html> <head> <title>Bienvenidos al Webber</title>
<!-- Meta caracteres -->
<meta name="robots" content="noindex, nofollow">
<meta name="keywords"
content="webber, procesador, processor, web, metainformacion, metadata">
<meta name="description" content="Webber es una herramienta para
desarrollo de entornos web">
<!-- Dublin Core -->
<META NAME="DC.Title" CONTENT="Bienvenidos al Webber">
<META NAME="DC.Creator" CONTENT="Webmaster de RedIRIS -
webmaster@rediris.es">
<META NAME="DC.Subject"
CONTENT="webber, procesador, processor, web, metainformacion, metadata">
<META NAME="DC.Description" CONTENT="Webber es una herramienta para
desarrollo de entornos web">
<META NAME="DC.Publisher" CONTENT="CSIC/RedIRIS. España">
<META NAME="DC.Date" CONTENT="2001-02-15">
<META NAME="DC.Language" CONTENT="es">
```

Dentro de las aplicaciones comerciales específicas destacan las de HiSoftware (*Hiawatha Island Software*) <<http://www.hisoftware.com>; <http://www.taggen.com>>, una de las empresas que mayor apuesta ha hecho por los metadatos en el desarrollo de sus productos. Algunos de sus software están destinados a lo que denominamos "la promoción de la Web", tal es el caso por ejemplo de **Hi-Visibility**¹⁶³ que instala

¹⁶² Ejemplo tomado del código fuente de la propia página de Webber, *Ibid.* Otras páginas constituidas con esta aplicación, donde se puede comprobar su código fuente, totalmente homogéneo con el de este ejemplo, son: las propias páginas de RedIris <<http://www.rediris.es>> y del CICA <<http://www.cica.es>>; las páginas del X Congreso Español sobre Tecnologías y Lógica Fuzzy <<http://www.imse.cnm.es/Estylf2000>>; y las páginas de Arraquis (Univ. de Murcia) <<http://arraquis.dif.um.es>>.

¹⁶³ Programa-versión: Hi-Visibility™ 3
Homepage: <http://www.hisoftware.com/hivisibility.htm>
Tipo: Demo (15 días)
Precio: \$ 49.99
Plataforma: Win 95/98/NT/2000/ME
Nombre (tamaño) del fichero: visibility.exe (4629 Kb)

[cont.]

automáticamente el *Engine Definition Builder* y que constituye una herramienta de gran utilidad para aumentar de forma certera y programada la visibilidad de un sitio Web en los motores de búsqueda genéricos. Estrictamente en relación con el proceso de creación de metadatos podemos destacar: **TagGen**¹⁶⁴, prestando especial atención a TagGen Office que integra la mayor parte de las utilidades para gestión de metadatos de Hisoftware, en una *suite* de programas.

TagGen es la principal aplicación de Hisoftware para la creación y gestión de metainformación. Se comercializa en diversas modalidades: TagGen DC Editor, TagGen para FrontPage, TagGen 4.5 para Allaire ColdFusion y HomeSite y TagGen Office que recopila todos los anteriores además de otras utilidades. Todas estas versiones del programa contemplan una especificidad de aplicación, sin embargo comparten formas de trabajo, aplicaciones complementarias y la interfaz. Analizaremos en detalle la aplicación TagGen Office, ya que es integradora del resto,

Descarga: <http://www.hisoftware.com/downloads/visibility.exe>

Fecha de descarga: 14/07/2001 20:58

¹⁶⁴Programa-versión: TagGen®

Homepage: <http://www.hisoftware.com/TagGen.htm>

Tipo: Demo (15 días)

Precio: Según versiones.<<http://www.hisoftware.com/othord.htm>>, TagGen Office \$399,99

Plataforma: Win 95/98/NT/2000/ME

Nombres (tamaños) de los ficheros:

- TagGen DC Editor: taggendc.exe (6706 Kb)
Descarga: <http://www.hisoftware.com/downloads/taggendc.exe>
Fecha de descarga: 22/03/2001 21:48
- TagGen Frontpage: taggen4.exe (6521 Kb)
Descarga: <http://www.hisoftware.com/downloads/taggen4.exe>
Fecha de descarga: 26/03/2001 15:59
- TagGen 4.5 for Allaire ColdFusion and HomeSite: tgallaire.exe (6740 Kb)
Descarga: <http://www.hisoftware.com/downloads/tgallaire.exe>
Fecha de descarga: 16/04/2001 22:07
- TagGen Office: taggene.exe (11136 Kb)
Descarga: <http://www.hisoftware.com/downloads/taggene.exe>
Fecha de descarga: 22/03/01 21:20

Comentarios: Más información sobre el funcionamiento de TagGen Office o de sus aplicaciones individuales, en: *Hisoftware TagGen™ Office: User's Guide* [documento PDF, comprimido ZIP].

Concord: Hiawatha Island Software Company, rev. 21 de junio de 2001. Disponible en: <http://www.hisoftware.com/DocCentral/taggenepdf.zip> (consultado el 18 de julio de 2001).

aunque haremos referencia a las características individuales de algunas de las herramientas que componen dicha *suite*, y que también se comercializan de forma individual.

Una característica común a todas las modalidades comerciales de este software, es que al ejecutar el *setup* se instalan otras dos aplicaciones: File List Builder que permite crear la lista de ficheros que se van a usar (por defecto .htm y .html, a los que se pueden añadir otras extensiones, p. ej. .xml, .rdf, —Fig. 33—); y Profile Builder que sirve para la creación de perfiles (con la extensión .mdpf) que pueden importarse posteriormente del menú *profile* de TagGen. La figura 34 presenta un ejemplo de creación de uno de estos perfiles de metainformación (el perfil por defecto), al que se añaden o eliminan tantos campos (etiquetas <META> en realidad) como sea necesario.

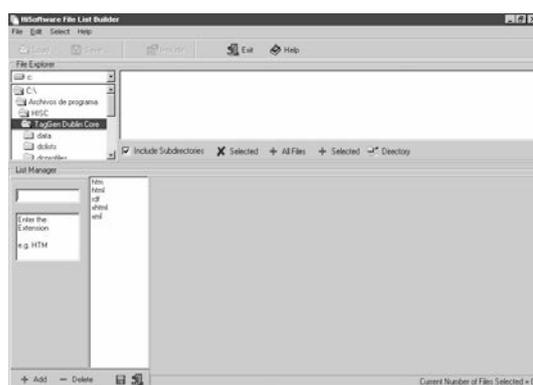


Fig. 33. Hisoftware File List Builder

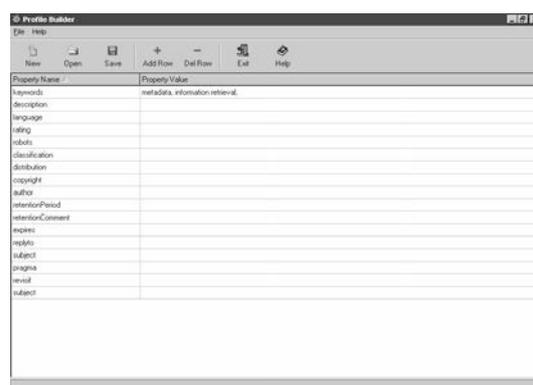


Fig. 34. Hisoftware Profile Builder

Como venimos anticipando, **TagGen Office** está compuesto por una *suite* de programas: (TagGen 4.5, TagGen 4.5 Dublin Core editor, TagGen 4.5 para FrontPage 2000 y 98, TagGen 4.5 para Allaire ColdFusion y HomeSite, MetaContents Builder, Hi-Verify, Hi-Verify para Microsoft FrontPage 2000 y 98) que recrean, desde nuestro punto de vista una edición y análisis de metainformación adecuada a las necesidades de cada usuario y/o tipo de información. Analizamos a continuación las prestaciones de cada una de ellas a fin de evaluar el rendimiento global del conjunto. En una primera aproximación a la complejidad y calidad de esta

herramienta, la tabla 10 presenta la estructura de directorios y los ficheros que almacena la instalación de TagGen Office.

Directorio	Ficheros
HISC\TagGen Office	Ficheros ejecutables de la aplicación y archivos de ayuda
HISC\TagGen Office\data	Listados controlados de valores (.dat)
HISC\TagGen Office\dclists	Ficheros DC
HISC\TagGen Office\dcprofiles	Perfiles DC (.mdfp)
HISC\TagGen Office\list temp	Ficheros temporales
HISC\TagGen Office\Lists	Listas de materia (.hstf, .txt)
HISC\TagGen Office\metacontents	Metacontents (aplicación)
HISC\TagGen Office\Profiles	Perfiles propios de TagGen (.mdpf)
HISC\TagGen Office\Save	Keywords (.kwf) y descripciones (.dsc) guardadas
HISC\TagGen Office\Schemas	Esquemas (.tgs)
HISC\TagGen Office\Hiverify	Hi-Verify (aplicación)
HISC\TagGen Office\Wbc	Ficheros relacionados con la actualización de Internet

Tabla 10. Estructura de directorios e información de TagGen Office

TagGen Professional (TagGen 4.6) permite incluir metadatos en las páginas Web trabajando bien en una unidad local o de red. Se trata de una aplicación muy flexible que se puede parametrizar a la medida del usuario. Algunas de las características más destacables de esta aplicación son:

- Edita y crea etiquetas, añadiendo o eliminando elementos de un esquema determinado de metainformación a través del *Custom Field Editor*.
- Tiene asociada una aplicación para construir una especie de "listas de autoridad" (*List Editor*) de tal forma que se pueda normalizar el contenido de algunas etiquetas (v.gr. *classification*, *rating*, *language*, etc.), que aparecerán como menús por defecto para completar el contenido durante el proceso de creación de metadatos.

- Distingue entre las metaetiquetas tipo <META NAME> y <META http-equiv> y dentro de estas últimas presenta el formulario para asignar al DLO el *Bulletin date* (utilizando un calendario para la selección) y el *Bulletin text* (seleccionado también de una lista editada *ad hoc*).
- Chequea la grafía, en inglés, tanto de las *keywords* como de la descripción; permite chequear los valores de las distintas etiquetas según unos valores predefinidos por el usuario (*Meta tag Analysis*).
- Tiene un completo fichero de ayuda tanto sobre el funcionamiento del programa y de todas sus prestaciones como sobre las etiquetas que soporta.

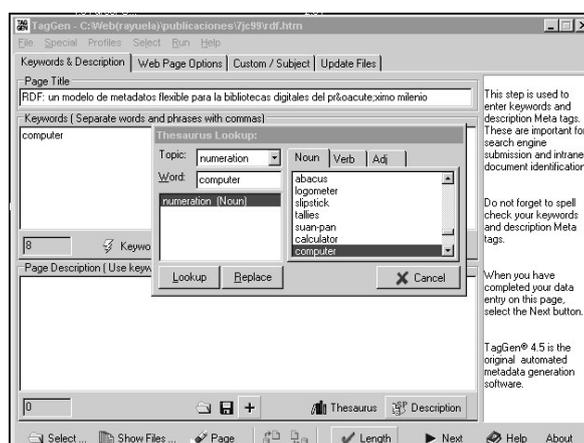


Fig. 35. TagGen 4.6 Professional

- Una de las características más destacables es que contempla el trabajo con Tesoros, tanto para la selección de descriptores como para la descripción¹⁶⁵ (Fig.35).

¹⁶⁵ Destacamos esta característica de TagGen Professional por la importancia que tiene a la hora de crear bibliotecas digitales temáticas donde la terminología debe de estar previamente controlada. Esta función de TagGen es al proceso de creación de metadatos ("catalogación de DLOs") lo que el módulo de Tesoro es a un SIGB convencional dirigido a la organización y gestión de publicaciones impresas o tradicionales. Trataremos someramente este complejo tema de la normalización del vocabulario y de los metadatos orientados a materias en el capítulo 6 (Vid. 6.4).

- Al igual que otras herramientas que hemos analizado anteriormente, puede realizar la extracción automática de descriptores (*Batch Keyword Generation*) permitiendo, previamente a su incorporación, la validación por parte del usuario. Para mejorar la indización automática tiene predefinida una lista de palabras vacías en inglés, pero también facilita la edición de listados con esta función.
- Tanto las palabras clave como las descripciones se exportan a ficheros .kwf, (en el caso de las *keywords*) y .dsc (en el caso de las descripciones), que posteriormente podrán importarse y reutilizarse en el caso de que estemos tratando un recurso de la misma temática.
- Además de todas las prestaciones que hemos resaltado en la creación de palabras clave, tiene su propio sistema para la asignación de materias (*Custom/Subject*), desde donde se puede crear una lista de materias jerárquica (con extensión .hstm) a través de la aplicación relacionada *Hisoftware Subject Tree Manager*, o bien visualizar algunas de las listas existentes¹⁶⁶.
- Permite al usuario crear sus propios esquemas (*schema BASE*) o utilizar un esquema predefinido de entre: el Dublin Core, el A-Core, el AGLS (*Australian Government Locator Service*), Publish, esquema específico sobre el estilo de publicación de los metadatos, GILS (*Global Information Locator Service*); IMS (*Instructional Management Systems*), el esquema propio de Microsoft (*MS Extended schema*) y WAGILS utilizado por el *Washington State Government Information Locator Service*.

¹⁶⁶ Además de la lista de materia jerárquica por defecto default.hstf, se pueden añadir a TagGen otras listas HSTF a partir del URL: http://www.hisoftware.com/taggen_files.htm, desde donde se pueden descargar distintos ficheros en formato comprimido (.zip) y pasarlos al directorio HISC\TagGen\Lists. *Washington GILS Subject Tree* (wagils.hstf), *RI GILS Subject Tree* (RI.hstf), *Utah GILS Subject Tree* (utah.hstf), *Illinois GILS Thesaurus y Topic Tree: Hierarchical* (states.hstf).

TagGen Dublin Core Editor, por su parte, facilita a los usuarios de TagGen Office trabajar y crear metaetiquetas según el modelo DC, utilizando tanto esquemas, como modificadores (*modifiers*) que permiten matizar más las etiquetas propias del Dublin Core a las necesidades concretas, y enlaces Web. Según la propia empresa¹⁶⁷, este programa se utiliza sobre todo por bibliotecarios, archiveros y grupos dedicados a la información gubernamental, con lo cual podemos esperar en él una mayor adecuación para la organización de páginas Web y su gestión en bibliotecas digitales o podemos explicarnos que tenga dentro de sus prestaciones una utilidad para convertir al formato DC, registros en MARC (Fig. 37).

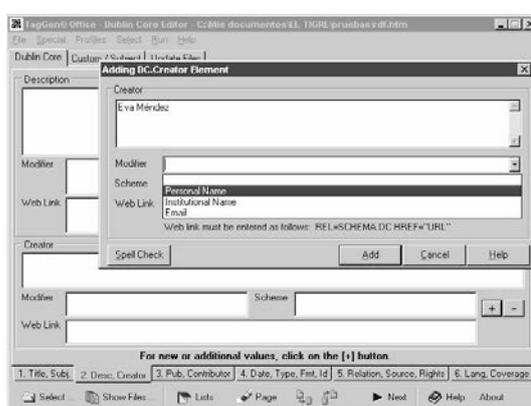


Fig. 36. TagGen Dublin Core Editor

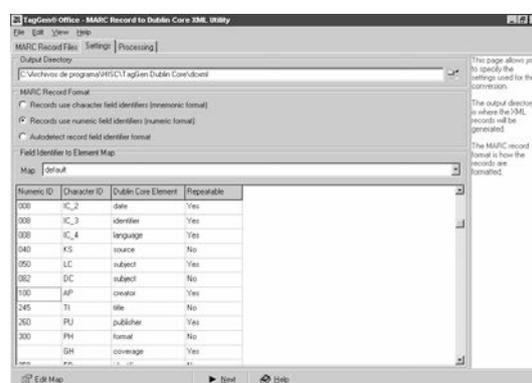


Fig. 37. TagGen Dublin Core Editor (conversión MARC)

Esta versión de TagGen para gestionar metadatos Dublin Core es, como muestran las imágenes (Figs. 36 y 37), una herramienta sencilla, completa y flexible para la creación de metainformación según este modelo. Permite crear los quince elementos del esquema DC, a través de seis secciones: 1) título y materia, 2) descripción y creador 3) editor y contribuidores 4) fecha, tipo, formato e identificador, 5) relación, fuente y derechos y 6) idioma y cobertura. Además de tener una interfaz semejante a TagGen Professional, también comparte sus prestaciones en

¹⁶⁷ Hisoftware TagGen™ Office, Op. cit., <http://www.hisoftware.com/DocCentral/taggenpdf.zip>, p. 13, 108.

cuanto a edición y modificación de perfiles y listados de autoridad. Sin embargo, en este caso resulta extraño que no contemple el concepto de tesauro para las materias, aunque sí se pueden seleccionar en virtud de los vocabularios que tiene almacenados¹⁶⁸. Otra de las particularidades que tiene la versión DC de TagGen, es la posibilidad de trabajar con etiquetas <META> de HTML, así como en modo XML. Esto añadido a la posibilidad de conversión de registros MARC a DC (Fig. 37), hacen de TangGen Dublin Core Edition una herramienta muy interesante para la creación y mantenimiento de metainformación, bien como utilidad complementaria en TagGen Office, bien como software aparte.

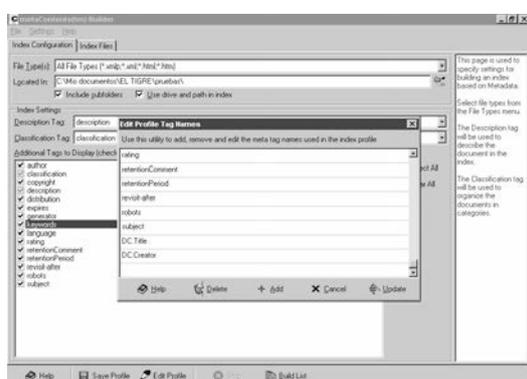


Fig. 38. Metacontents™ Builder

Metacontents Builder¹⁶⁹, incluida en TagGen Office, pero comercializada también como herramienta aparte, está destinada a la indización y creación de metainformación pero en este caso con la finalidad concreta de crear sitios Web basados en metadatos. Su funcionamiento no difiere de las herramientas que hemos analizado en el

apartado dedicado al software cliente para la creación de metaetiquetas (4.3.1.2) pero más que un editor es un visualizador o presentador de metainformación. Extrae los metadatos del documento o documentos de un directorio seleccionado, según una

¹⁶⁸ Además de esos vocabularios, se pueden crear otros o introducir Tesauros preexistentes en formato .hstf. *Cfr. sup.* Nota 166.

¹⁶⁹ Programa-versión: Metacontents Builder™ 2.0.1
 Homepage: <http://www.hisoftware.com/metaContents.htm>
 Tipo: Demo (15 días)
 Precio: \$ 49.99
 Plataforma: Win 95/98/NT/2000/ME
 Nombre (tamaño) del fichero: metacb.exe (4230 Kb)
 Descarga: <http://www.hisoftware.com/downloads/metacb.exe>
 Fecha de descarga: 22/03/2001 22:10

serie de etiquetas (por defecto: author, classification, copyright, description, distribution, expires, generator, keywords, language, rating, retentionComment, retentionPeriod, revisit-after, robots, subject, aunque se pueden añadir tantas etiquetas como sea preciso¹⁷⁰); de esa forma facilita la creación de índices de contenido según esta metainformación que extrae y los visualiza en forma de página Web, o lo que Hisoftware denomina "navegación tipo portal" o "tipo Yahoo" (*Portal style navigation*¹⁷¹, *Yahoo navigation*). En realidad, podría decirse que es una aplicación para crear sitios Web organizados según un criterio de clasificación temático, al estilo de los *pathfinders* de CORC, pero partiendo de la estructura de metadatos de las páginas. En este sentido puede resultar muy útil para la organización y edición fuera de línea (p. ej. CD-Rom) de información en HTML, y en cualquier caso, un complemento interesante al resto de aplicaciones de TagGen Office. Además de la flexibilidad que tiene en la selección de metainformación para visualizar y en la forma de visualización, trabaja con documentos HTML, XML e incluso paquetes de metadatos en XML (XMLP) generados con **MetaPackager**¹⁷².

¹⁷⁰ En la Fig. 38 se muestra el proceso de introducción de nuevas etiquetas en el perfil de metadatos, concretamente en este caso, DC.Title y DC.Creator. Aunque más controlada la selección de metadatos por el usuario a través de estos perfiles, los sitios Web que se crean a partir de Metacontents Builder se asemejan a la creación de *pathfinders* en CORC. Cfr. *supr.*, Vid. Glosario, *pathfinders*.

¹⁷¹ Este concepto aparece cada vez con más frecuencia en el ámbito de los metadatos, con el término *portal style navigation*, como en Metabuilder <<http://www.hisoftware.com/metaContents.htm>> o con otras denominaciones como *portal-style interfaces (such as Yahoo)*, como alude a esta forma de estructurar la información, la empresa Watchfire creadora de Metabot (Vid. *infr.*). *Metatags: They're not Just... Op. cit.*, <http://www.watchfire.com/resources/metatagswhite.pdf>, p. 4.

¹⁷² Programa-versión: MetaPackager™ 2.0

Homepage: <http://www.hisoftware.com/metaPackager.htm>

Tipo: Demo (15 días), requiere registro y password enviada por correo electrónico.

Plataforma: Win 95/98/NT/2000/ME

Nombre (tamaño) del fichero: mpro.exe (9960 Kb)

Descarga: <http://www.hisoftware.com/downloads/mpro2.exe>

Fecha de descarga: 30/06/2001 06:45

Comentarios: permite utilizar metadatos para indizar y recuperar todo tipo de objetos digitales, independientemente del tipo de fichero, y sobre todo, está destinado a aquellos tipos que no tienen la

[cont.]

El último componente de software de la *suite* de TagGen Office que, al igual que los anteriores también se comercializa aparte, es Hi-Verify. Este software podría definirse como un analizador del valor de las etiquetas, que evalúa y prepara las páginas con metainformación tanto para la promoción Web como para la recuperación. **Hi-Verify**¹⁷³ es una aplicación complementaria de TagGen muy interesante, pues demuestra que no sólo es importante asignar metadatos a los recursos Web, sino que este proceso requiere una comprobación de la calidad de la metainformación para que ésta sea eficaz. Entre las funciones de este software podemos destacar la verificación de descriptores. Realiza un chequeo de repeticiones (*spamming*¹⁷⁴), analiza las palabras clave utilizadas en la etiqueta *description*, chequea también el uso de las *keywords* utilizadas en las etiquetas <META> con las utilizadas en el <BODY> del documento¹⁷⁵, y hace una comprobación de aquellas etiquetas que se han definido como requeridas u obligatorias. En el caso de que haya alguna anomalía en alguna de estas comprobaciones dará un mensaje de error en el informe que genera automáticamente y que se visualiza también en forma de página Web.

La forma de trabajar de Hi-Verify es muy sencilla: partiendo de un listado determinado de ficheros HTML, archivados bien en local o en la Red, realiza un

capacidad o la facilidad de embeber los metadatos de forma convencional (archivos PDF, documentos de MS Office, imágenes, sonidos, mapas, etc.), almacenando la metainformación en XML.

¹⁷³ Programa-versión: Hi-Verify™ 3.0

Homepage: <http://www.hisoftware.com/verify.htm>

Tipo: Demo (15 días)

Precio: \$49,99

Plataforma: Win 95/98/NT/2000/ME

Nombre (tamaño) del fichero: verify.exe (4364 Kb)

Descarga: <http://www.hisoftware.com/downloads/verify.exe>

Fecha de descarga: 23/03/2001 15:47

¹⁷⁴ Para una definición exacta y completa de este término y de su uso, *Vid.* Glosario, *spam*, *spamming*.

¹⁷⁵ Esta comprobación es muy importante ya que si el motor de búsqueda no encuentra referencias a las palabras clave en el documento, otorgará una puntuación más baja a los valores de los metadatos y por tanto, menor relevancia a las páginas. En general, para la mayoría de los motores de búsqueda se recomienda usar cada una de las palabras clave utilizadas en las metaetiquetas al menos cinco veces.

análisis de la metainformación contenida en ellos, basándose en una serie de reglas seleccionadas por el usuario y denominadas *Rule base*.

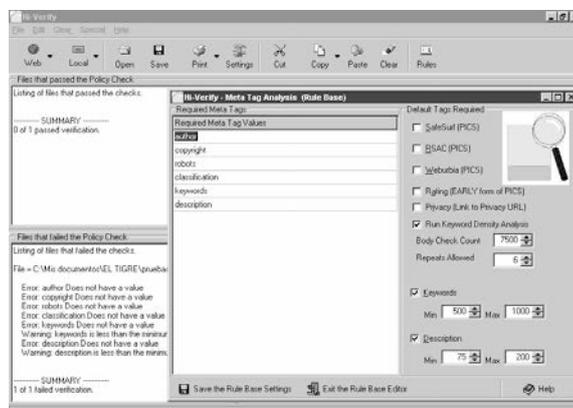


Fig. 39. Hi-Verify™ 3.0

La apuesta que ha hecho Hisoftware en el reconocimiento de los metadatos para el desarrollo de sus productos les ha llevado también a desarrollar otra *suite* de programas denominada **JITI Builder Office**¹⁷⁶, cuyo interés reside en que facilita la gestión de las propiedades de los documentos de Microsoft Office. Los módulos que componen JITI son:

- JITI Builder, para asignar metadatos en los documentos de Microsoft Word, Excel y PowerPoint. (La Fig. 40 muestra como ejemplo, el proceso de asignación de metadatos para el documento de Word que constituye el capítulo 4 de esta tesis doctoral).

¹⁷⁶ Programa-versión: JITI Builder™ Office 1.5

Homepage: <http://www.hisoftware.com/Franklin/JITIBLD.htm>

Tipo: Demo (15 días)

Precio: \$295.00

Plataforma: Win 95/98/NT/2000/ME

Nombre (tamaño) del fichero: joffice.exe (6692 Kb)

Descarga: <http://www.hisoftware.com/downloads/joffice.exe>

Fecha de descarga: 14/04/2001 02:45

- JITI Verify, similar a Verify en TagGen, analiza y verifica los metadatos tanto a nivel local como Intranet o ficheros compartidos antes de publicarlos en la Red.
- MetaSearcher, que sirve para buscar documentos de Office basándose para ello en los metadatos asignados y en el tipo de documentos.
- MetaContents Builder, al igual que el módulo del mismo nombre en TagGen, sirve para crear índices en HTML basándose para ello en la estructura de metadatos asignada.

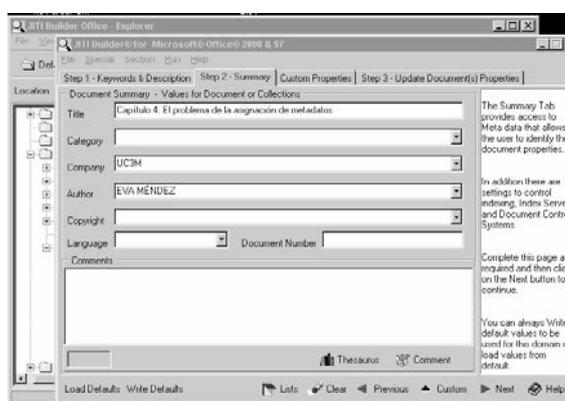


Fig. 40. JITI Builder® for Microsoft Office 2000 & 97

JITI, también instala las dos herramientas básicas de Hisoftware: File List Builder, (Fig. 33) donde las extensiones por defecto son las propias de los documentos ofimáticos del Microsoft Office (.doc, .ppt, .xls) a las que podemos añadir otras extensiones de ficheros; y Profile Builder (Fig. 34) que sirve, al igual que en el resto de las aplicaciones de HiSoftware, para crear perfiles *ad hoc*. Esta herramienta nos parece especialmente útil para la gestión documental y para incrementar las posibilidades de recuperación de información basada en metadatos en un contexto finito de información como son por ejemplo, las Intranets corporativas¹⁷⁷, donde se producen diariamente multitud de documentos ofimáticos,

¹⁷⁷ A pesar de que en este trabajo defendemos la tesis de la utilidad y aplicabilidad de los metadatos en bibliotecas digitales, partimos de una afirmación más genérica de la utilidad de esta tecnología en [cont.]

textuales, hojas de cálculo, etc. cuya descripción y localización puede incrementar el potencial de control sobre el conocimiento tácito de una organización. Además podemos extender o prever la utilidad de JITI al contexto genérico y global de la WWW. Entre la riqueza de formatos y de tipos de DLOs que podemos encontrar en la Red, cada vez es más frecuente que existan presentaciones .ppt u otros documentos ofimáticos realizados con alguna de las aplicaciones de uso estandarizado y generalizado de Microsoft. Al ser ficheros realizados con un formato propietario y no estar basados en el mercado, corren el riesgo de formar parte, al menos para algunos buscadores, de lo que se denomina *Internet invisible*¹⁷⁸ de no tener una descripción de sus propiedades.

Otro de los software más interesantes para la creación y visualización de metadatos es **Metabrowser**¹⁷⁹, una herramienta de software en constante evolución,

contextos finitos de información. Por ello es especialmente significativo hablar aquí, a tenor de JITI, de las Intranets, que pueden ser otro paradigma de aplicación de los metadatos según la argumentación que aportamos de esta investigación, ya que su naturaleza y crecimiento permiten un control del conocimiento corporativo basado en metainformación.

¹⁷⁸ Se denomina *Internet invisible* al conjunto de recursos que no pueden ser indizados de forma estructural y genérica por los motores de búsqueda. Además de las distintas bases de datos, bibliográficas o no, también pertenecen a esta categoría de lo que Isidro Aguillo llama "infranet", los documentos en formatos no indizables, esto es, aquellos documentos contruidos con lenguajes de descripción de páginas (Acrobat, Postcript) y también los documentos ofimáticos realizados con un software propietario como es el caso de los ficheros de Word, Excel y PowerPoint. Vid. Isidro Aguillo. *Internet invisible o Infranet: definición, clasificación y evaluación*. En: *Jornadas Españolas de Documentación* (7. 2000. Bilbao). *Op. cit.*, p. 249-269. Aunque posteriormente, los principales motores de búsqueda han incluido entre los formatos susceptibles de indización todos los tipos citados. Para la búsqueda de información de los documentos no visibles, en sentido estricto (bases de datos, etc.), Vid. <http://www.internetinvisible.com>

¹⁷⁹ Programa-versión: Metabrowser 1.2.294

Homepage: <http://metabrowser.spirit.net.au>

Tipo: Demo (30 días); Navegador (*freeware*)

Precio: Desde \$AU695 (\$US 351,14)

Plataforma/requisitos: Win 95/98/NT/2000/ME, mínimo 32Mb RAM

Nombre (tamaño) del fichero: metabrowser.exe (2044 Kb)

Descarga: <http://metabrowser.spirit.net.au/download/metabrowser.exe>

Fecha de descarga: 19/06/01 06:12

Comentarios: La división de Servicios de Metadatos de la Utah State Library, tiene su propia versión adaptada de Metabrowser. Información disponible en: <http://www.utah.org/gils/tutorial.htm>

desarrollada por la empresa australiana Spirit Consulting para la visualización y creación de metadatos y canales RSS, en una sintaxis tanto RDF como HTML, y con una gran cantidad de posibilidades de gestión de esquemas y vocabularios para la asignación de metadatos.

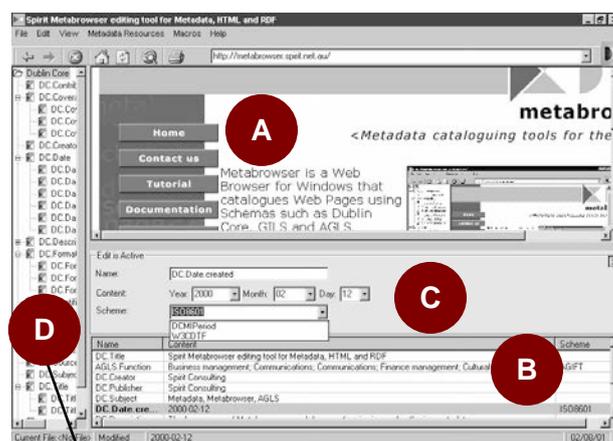


Fig. 41. Metabrowser

Vid. Fig. 42

Presenta una sencilla interfaz compuesta por cuatro paneles de utilidades:

- A) Un navegador Web, con funcionalidades distintas a los navegadores convencionales, tipo IE Explorer o Netscape, dirigidas a la visualización y gestión de metainformación.
- B) Un visualizador de los metadatos asociados a una determinada página mostrada.
- C) Un panel de edición de metadatos. Esta división de ventana en Metabrowser se abrirá también para la modificación de alguno de los campos de metainformación visualizados en el panel b.
- D) Un visualizador de esquemas (*schemas* y *schemes*) de metadatos en forma jerárquica (Vid. Fig. 42).

La figura 41 muestra estas cuatro funciones de la aplicación y la 42, muestra individualmente el visualizador jerárquico. En la demo que estamos evaluando, editor de metadatos y los catálogos XML están limitados a 30 días, sin embargo el navegador es *freeware*. Las nuevas características de la versión 1.2.xxx que hemos

utilizado¹⁸⁰ incluyen soporte para múltiples dominios y formatos de metadatos <<http://metabrowser.spirit.net.au/tutorial.html>> para XML/RDF, y un fichero de ayuda y unas características de interfaz de usuario mejoradas.

Como navegador Web, Metabrowser tiene las mismas prestaciones que cualquier otro agente de usuario (AU) (p. ej.: barra de herramientas, preferencias, posibilidad de abrir varios navegadores, abrir ficheros remotos o locales, perfecta integración con el sistema operativo de Windows, y un menú contextual —botón derecho del ratón— idéntico e integrado con IE); pero además permite visualizar los metadatos incluidos en la página Web. El tema de la visualización de los metadatos, salvo en algunas excepciones¹⁸¹, se ha pasado bastante por alto en la investigación sobre bibliotecas digitales, no obstante alberga una problemática muy interesante. Mientras que las interfaces gráficas de usuario de los AU más conocidos (v. gr. Netscape Navigator, Microsoft IE, Opera, etc.) *están orientados a una navegación visual, existe todavía una gran escasez de aplicaciones que permitan una visualización gráfica del contenido*¹⁸². Metabrowser sin embargo, está adaptado para visualizar al mismo tiempo tanto la información Web en sí misma, como el sustituto (*subrogate*¹⁸³) del DLO, donde se pueden además, editar y modificar los metadatos.

¹⁸⁰ Concretamente hemos evaluado la versión 1.2.294 (*Ibid.*)

¹⁸¹ Beagle es uno de los pocos autores que han investigado las posibilidades de las técnicas de visualización de metadatos y de las representaciones informativas (*surrogates*) basadas en ellos, aludiendo a las nuevas sinergias entre interrogación, navegación y compartir información que puede generar la visualización de metadatos en la utilización de la información digital. Donald Beagle. Visualization of Metadata. *Information Technology and Library*, December 1999, vol. 18, n° 4, p. 192-199. También disponible en: http://www.lita.org/ital/1804_beagle.html

¹⁸² *Ibid.*, p. 192.

¹⁸³ *Vid.* Glosario.

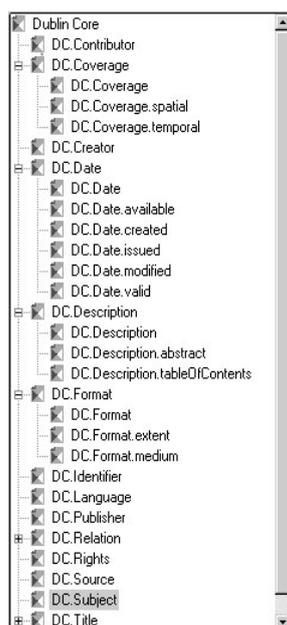


Fig. 42. Metabrowser (Tree Viewer – DC Schema)

Como herramienta para la creación de metadatos, Metabrowser permite: visualizar en el navegador la página (local o remota) a la que se quieren asignar metadatos, una plantilla para rellenar los campos de metainformación, así como ampliar la plantilla según las necesidades del usuario, y asociar ésta finalmente, a una tecla de función.

Además la herramienta *Tree View* que incluye Metabrowser, asiste en la creación de metadatos facilitando tanto la visualización como la edición de listas en forma jerárquica de las estructuras de datos que se utilizarán en el proceso de asignación (Fig. 42), bien sean de elementos del esquema de metadatos utilizado

(*schemas*¹⁸⁴) o listas controladas para el contenido (*schemes*¹⁸⁵) que servirán indiscutiblemente para homogeneizar la entrada de metadatos, y también índices de

¹⁸⁴ Vid. Glosario: Diferenciación entre *schema* y *scheme*. Los esquemas (*schema*) de metadatos con los que trabaja *a priori* Metabrowser son: DC, AGLS (2 versiones), BEP (*Business Entry Point*), EdNA (*Educational Network Australia*), GILS-Utah, y el modelo de metadatos del Health Institute del Gobierno Australiano, sin que ello obste para que se puedan desarrollar, como hemos dicho, otros *schemas* o adaptaciones de ellos.

¹⁸⁵ Vid. Glosario: Diferenciación entre *schema* y *scheme*. Los esquemas (*scheme*) o vocabularios para controlar la entrada de metadatos en Metabrowser son, *a priori*, y con relación a cada uno de los *schemas* que soporta:

- Con el AGLS: AGIFT (*Australian Governments' Interactiva Functional Thesaurus*), Aust X500 y IMT (*Internet Media Types*), para la etiqueta relativa al formato DC.Format.
- Con BEP: BEP.industry, BEP.jurisdiction, BEP.location, BEP.topic, BEP.type.
- Con el Dublin Core: DCMIType, IMT (DC.Format) y RFC1766 para la identificación del idioma.
- Con el modelo EdNA: edna.audience, edna.coverage, edna.curriculum, edna.directory, edna.document, edna-event, edna.indexing, edna.KLA, edna.sector, edna.userlevel, DCMIType, IMT (DC.Format) y RFC1766.
- Con GILS-Utah: Copyright, Departments, Doc Type, Language, Medium, Revisit After, Robots y Subjects.

materia (*subject index*) y para lo que se denomina *Harvest Control Lists* (HCL¹⁸⁶) que funcionan como los marcadores en los navegadores convencionales generando índices de referencia de metadatos, y que en el caso de Metabrowser están diseñados para publicarse en la Web y que cualquiera pueda acceder a ellos. En síntesis, esta utilidad (*Tree Viewer*) posibilita el desarrollo de un sistema más coherente de metainformación.

La excelencia del sistema de control de metadatos que tiene Metabrowser se debe a un sistema de macros que sirven para definir, a través de un protocolo propietario `mtp://` (Metadata Transaction Protocol), varias funciones, desde la creación de nuevos *schemas* y *schemas*, su visualización en el *Tree Viewer*, o el envío de metainformación a un CGI externo para facilitar la gestión de bases de metadatos externas. Todas las macros que gestiona este programa tienen la siguiente sintaxis:

```
mtp://field=[fieldname]{&content=[content]}
{&scheme=[scheme]}.fld
```

Así, por ejemplo, para definir la opción del *schema* que utiliza el DC para la fecha puede ser la norma ISO 8601, Metabrowser genera la siguiente macro:

```
mtp://field=DC.Date&scheme=ISO8601.fld
```

O `mtp://url=metabrowser.spirit.net.au/schemes/dc.mbs`, para indicarle al navegador que cargue el *schema* completo del DC en el visualizador jerárquico (*Tree Viewer*).

Otra completa aplicación para la creación y manejo de metadatos es **Metabot Pro**¹⁸⁷ desarrollado por la empresa Watchfire (antes Tetranet), dentro de su acervo de

¹⁸⁶ Vid. Glosario, Fig. 69.

¹⁸⁷ Programa-versión: Watchfire Metabot Professional 3.51.0.546
Homepage: <http://www.watchfire.com/solutions/metadata.asp>
Tipo: Demo (30 días). Requiere registro
Precio: \$495

soluciones para el acceso y la búsqueda de información, principalmente en el ámbito de la información empresarial y del comercio electrónico¹⁸⁸.

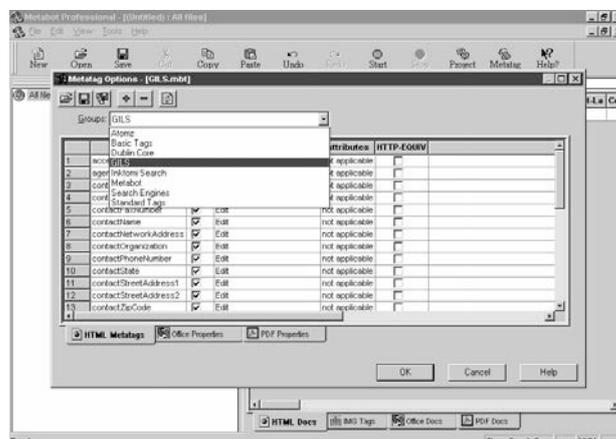


Fig. 43. Metabot Professional

Metabot es una herramienta muy completa para la gestión de metadatos. Permite más de 50 metaetiquetas¹⁸⁹ correspondientes a ocho modelos, que van desde formatos más o menos consolidados (como las 31 etiquetas de GILS para describir la información gubernamental o los quince atributos del Dublin Core sin calificadores¹⁹⁰), a conjuntos de etiquetas dirigidos a los sistemas de búsqueda (esto es, Description, Keywords, Refresh (http-equiv), Robots y

Plataforma/requisitos: Win 95/98/NT/2000/ME, mínimo 64Mb RAM (128Mb recomendado)

Nombre (tamaño) del fichero: metabot.exe (9692 Kb)

Descarga: URL personalizada por correo electrónico después del registro:

<http://www.watchfire.com/callhome/downloadcheck.asp?email=emendez@bib.uc3m.es&product=metadata> (7 días para descargar el software)

Fecha de descarga: 21/06/01 04:07

¹⁸⁸ Watchfire Enterprise Solutions: <http://www.watchfire.com/solutions/wes>

¹⁸⁹ Un listado de las cuarenta y tres metaetiquetas que permite para la descripción de ficheros HTML se puede encontrar en: <http://www.watchfire.com/solutions/metadata/meta-tags.asp> (aunque en la versión del programa que hemos evaluado, son 44 las etiquetas, las 43 que se describen aquí, más la etiqueta title). Esto no obsta para que se puedan generar otras tantas, como hemos dicho, y que utilice propiedades específicas para los ficheros .doc, xls, ppt, gif, jpg y .pdf.

¹⁹⁰ Formato DC *unqualified*, se conoce así desde la cuarta reunión del Dublin Core en Canberra (1997) al conjunto de los 15 elementos que conforman el modelo DC, sin incluirle los tres atributos calificadores SUBELEMENT, SCHEME y LANG. Vid. 5.2.1.2.

Title) o a un sistema de búsqueda en particular como AtomZ¹⁹¹ o Inktomi <<http://www.inktom.com>>, los atributos que puede llevar el elemento <META> —tanto con el atributo `name`, como con el atributo `http-equiv`— en HTML (*Standard Tags*) o las denominadas por Metabot *etiquetas básicas* (autor, título, palabras clave y descripción) (Fig. 43). Además de todas estas posibilidades y de permitir modificar estas estructuras o esquemas de metainformación, añadiendo o quitando etiquetas, se pueden crear y guardar modelos o plantillas *ad hoc* bien utilizando como base alguno de estos formatos, bien creando uno totalmente nuevo.

Metabot, al igual que veíamos en JITI de Hisoftware, es una herramienta de gestión de metadatos muy dirigida al ámbito de la información empresarial y del comercio electrónico. Por ello tiene un sistema de asignación de metainformación para los documentos de Microsoft Office (.doc, .ppt, .xls) a través de lo que denomina *Office properties* y para los documentos pdf (*PDF properties*) y que no son más que las propiedades que permite generar cada uno de los programas propietarios de dichos formatos. Además, aunque no sigue ningún modelo de metainformación formal para las imágenes¹⁹², sí identifica las siguientes características: nombre del fichero, y los atributos `src`, `alt`, `width`, `height`, posibilitando de este forma una descripción mínima de estos archivos no textuales¹⁹³. Toda esta metainformación

¹⁹¹ AtomZ <<http://www.atomz.com>> es un motor de búsqueda que, junto a htDig <<http://htdig.sourceforge.net>> y Mamma <<http://www.mamma.com/acb/>>, es uno de los más utilizados para hacer búsquedas internas en un sitio Web. La propia empresa Watchfire utiliza AtomZ <<http://www.watchfire.com/search/search.asp>> como buscador interno dentro de su sitio. Las etiquetas que contempla este modelo, son obviamente las que soporta este buscador: `description`, `keywords`, `last-modified`, `robots`, `target`, `title`. No obstante, AtomZ ha hecho una fuerte apuesta últimamente por los metadatos para la recuperación de información en sitios Web. Cfr. 7.2.

¹⁹² En el apartado 5.2.2.2. se mencionarán algunos de los esquemas de metainformación constituidos especialmente para la descripción y recuperación de imágenes, tales como VRA ó DIG35.

¹⁹³ Analiza los ficheros de imagen (GIF y JPEG) y sugiere texto alternativo (atributo ALT) para estos ficheros no textuales. Esta característica de Metabot es muy interesante en términos de accesibilidad ya que proporciona información textual para elementos gráficos y permite no perder esa información [cont.]

que maneja se guarda en *metabot project files* (.mbp) que pueden volver a editarse, permitiendo así mismo el trabajo en modo *batch*.

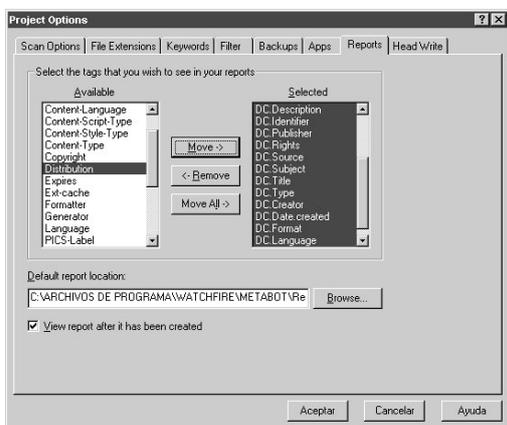


Fig. 44. Metabot Professional. Selección de etiquetas para informe.

Se trata pues de una herramienta sencilla, flexible e integrada. Su sencillez se manifiesta en el completo fichero de ayuda con el que cuenta, así como en los cuadros de diálogo que dan al usuario consejos para una mejor utilización y manejo del programa, o en la facilidad de una función (*wizard*) para guiar a los menos familiarizados, en el proceso de creación de metainformación a ficheros individuales. Es flexible porque permite

varias opciones de configuración tanto de las características de los proyectos (mbp) con los que trabaja, como de la propia interfaz que, a diferencia de otros software que basan su trabajo en plantillas o *templates*, Metabot trabaja con una especie de hojas de cálculo que facilitan las funciones de creación, eliminación o adición de metadatos. Además funciona tanto en local como vía FTP, de tal forma que los ficheros que se editan se pueden modificar y transferir inmediatamente. En términos de integración debemos hablar, por un lado, de la integración con otras herramientas de software para la edición de páginas Web, tales como Dreamweaver 3.0, Homesite o Frontpage, y por otro de la propia integración de funciones de Metabot.

Si bien sus funciones están perfectamente integradas para realizar una gestión global de metadatos, algunas de ellas pueden utilizarse individualmente con mucha rentabilidad. Tal es el caso de la generación de informes, donde se seleccionan las

ante el uso de navegadores no gráficos, utilizados generalmente por discapacitados. Vid. WCAG. World Wide Web Consortium. *Web Content Accessibility...* Op. cit., <http://www.w3.org/TR/WCAG10>

etiquetas que se quieren analizar (la Fig. 44 presenta etiquetas del DC) y genera un informe inmediato de su uso¹⁹⁴.

A pesar de las particularidades de cada uno de estos software, al mismo nivel que TagGen, Metabrowser o Metabot, debemos destacar otro denominado MetaManage¹⁹⁵ de la empresa, también australiana como en el caso de Metabrowser, Edime <www.edime.com.au>. Se trata de un cliente orientado a objetos que permite, entre otras cosas, la creación de sitios Web normalizados. Es una herramienta muy completa de la que podemos destacar las siguientes características:

- Al ser una herramienta orientada a objetos simplifica la configuración y gestión tanto de sitios Web, como de Intranets. Trabaja asimismo a nivel multiusuario de tal forma que se pueden gestionar metadatos de manera integral por varios creadores a la vez.

¹⁹⁴ Aunque como mencionamos en el apartado dedicado al marco y objetivos de esta investigación (2.1), excluimos de ella cualquier análisis puntual sobre el uso de metainformación, sí destacamos la metodología informática para realizar estos estudios. *Cfr.* Capítulo 2, notas 3 y 4.

¹⁹⁵ Programa-versión: Metamanage 2.5.15

Homepage: <http://www.metamanage.com>

Tipo: Demo (14 días). Requiere registro y es una versión limitada

Limitaciones <<http://www.metamanage.com/demo.htm>>: no incluye el *plug-in* de envío por correo electrónico, está limitada a la configuración de un solo sitio y la profundidad jerárquica está reducida a 5).

Precio: Desde \$495 (versión profesional: \$1600)

Plataforma/requisitos: Win 95/98/NT

Nombre (tamaño) del fichero: SetupMetaM2515.exe (2075 Kb)

Descarga: <http://cgi.edime.com/metamanage/download/download.cfm> (previa confirmación por correo electrónico de *login* y *password*)

Fecha de descarga: 17/05/01 03:36

Comentarios: El carácter mnemotécnico que se confiere a los nombres de las aplicaciones de metadatos (que, como hemos visto suelen incluir en prefijo *meta-*) puede llevar a confusión entre las mismas. Así, por ejemplo en este caso, debemos puntualizar la diferencia entre esta aplicación cliente para la gestión de metadatos con la plantilla para la edición metaetiquetas *Metamanager Site Owner* (Figs. 9 y 10) que citamos en el apartado dedicado a las herramientas para la creación de metaetiquetas. *Cfr.* 4.3.1.1.

- Crear fácilmente metadatos según los modelos AGLS y DC a un nivel de detalle de subelementos o calificadores muy preciso (Fig. 45). Sin que ello obste para insertar las metaetiquetas que reconocen algunos motores de búsqueda (*keyword* y *description*) para hacer los sitios más visibles en la generalidad de la WWW, a través de buscadores y portales de información genéricos.



Fig. 45. Metamanage

- Distingue entre estándares de estructura (*schemas*, como AGLS y DC) y estándares de contenido (*schemes* como por ejemplo RFC1766 para la etiqueta DC.Language, o tesauros para la asignación de materias) para los valores de los metadatos, y en ambos casos se pueden modificar y ajustar a las necesidades del usuario o del sistema de información.
- Una vez asignados todos los metadatos permite hacer operaciones globales sobre un conjunto de ficheros y/o directorios (búsqueda, cambio, borrado y ordenación de metadatos). Incluso facilita la interoperabilidad sincronizando etiquetas con el mismo contenido semántico (v. gr. <TITLE> y DC.Title).
- Una de las funciones más características que tiene Metamanage es la posibilidad de crear informes a los desarrolladores de sitios Web sobre cuando caduca el contenido o cuando requieren revisión las páginas.
- Además de su aplicación a la gestión de metadatos permite mantener mapas de sitios Web, y aplicar y gestionar hojas de estilo (CSS).

Metamanage es también una herramienta flexible y con muchas posibilidades para planificar un proceso de creación y gestión de metadatos.

4.2.2.3 Creación de metadatos en SIG-DLOs

Hasta ahora, la mayoría de las herramientas que hemos analizado, aunque tengan una fácil integración con bases de datos, como p. ej. Metabrowser, o con otros componentes de software, más complejos, como en el caso de Metabot, son aplicaciones independientes (*stand-alone*) que se comercializan individualmente para la creación de metadatos. Sin embargo, no podemos dejar de mencionar aquellas herramientas que son parte integral o un módulo de todo un sistema para la gestión de información en entornos WWW. A este tipo de software podríamos denominarlos, siguiendo las denominaciones más usadas, sistemas de gestión de documentos electrónicos¹⁹⁶ (sistemas GED), sistemas integrados para la gestión de la información digital, e incluso, de una forma más pretenciosa, como se autodenominan algunos de ellos, "software para la gestión del conocimiento". Sin embargo, por coherencia con toda la teoría que estamos desarrollando en torno al objeto de procesamiento y al fundamento de los metadatos, hemos decidido denominarlos Sistemas Integrados para la Gestión de Documentos entendidos como Objetos de información (**SIG-DLOs**), basándonos para ello de la denominación, en este caso más consolidada, que recibe el software para organizar y constituir bibliotecas convencionales, nos referimos sin duda a los Sistemas Integrados de Gestión de Bibliotecas (SIGB). Siguiendo el paralelismo con los SIGB, también se podrían designar sencillamente SIGBD (Sistemas Integrados de Gestión de Bibliotecas Digitales) pero, como ya hemos apuntado, la gestión de recursos digitales (DLOs) es algo mucho más complejo que la concepción y denominación formal de una biblioteca digital. Por ello utilizamos SIG-DLOs, de forma que se contemple también su utilidad para la gestión

¹⁹⁶ *Electronic Document Management Systems (EDMS)*, así los denominan por ejemplo Bielawski & Boyle y Sprague. L. Bielawski & J. Boyle. *Electronic Document Management Systems: A User Centered Approach for Creating, Distributing, and Managing Online Publications*. Saddle River, NJ: Prentice Hall PTR, 1997; R. Sprague. *Electronic Document Management: Challenges and Opportunities for Information Systems Managers*. *Management Information Systems Quarterly*, 1995, vol. 19, nº 1, 29-49. Citados en: Lisa D. Murphy. *Op. cit.*, p. 54.

de otras colecciones (locales o distribuidas) de objetos digitales que no se constituyen como biblioteca virtual/digital o electrónica.

Los SIG-DLOs de manera genérica son, en la mayor parte de los casos, herramientas comerciales que se han desarrollado para la creación de grandes *data warehouses*¹⁹⁷ en el contexto de la gestión de la información empresarial¹⁹⁸ y del comercio electrónico, pero ello no obsta para que puedan utilizarse también para gestionar colecciones digitales, o bien, son la respuesta de algunos (pocos) SIGB a la demanda de gestión de bibliotecas digitales o de la integración de recursos electrónicos en catálogos de bibliotecas tradicionales.

Dentro del primer grupo, herramientas para el desarrollo de *data warehouses* en entornos empresariales y para el *e-commerce*, podemos destacar, por ejemplo:

- **Metamatrix** <<http://www.metamatrix.com>> que, según describen ellos mismos¹⁹⁹, *proporciona una plataforma escalable para el comercio electrónico*. Se trata de una aplicación integral que sirve para almacenar y gestionar metadatos a nivel empresarial, proporcionando a los desarrolladores del sistema de información una visión global del contenido de los datos de toda la empresa. Lo que hace Metamatrix en relación con los metadatos es un almacenamiento centralizado de

¹⁹⁷ Vid. Glosario.

¹⁹⁸ Una visión muy interesante de los metadatos en el ámbito de la gestión de la información empresarial la encontramos en la obra de Toser, donde distingue tres niveles de la gestión de la información empresarial (estratégico, táctico y operacional) y describe un modelo para el tratamiento de dicha información basado en metadatos para pasar de una mera gestión, a un control de la información empresarial. Guy Tozer. *Metadata Management for Information Control and Business Success*. Boston, London: Artech House, 1999, 318 p.

¹⁹⁹ Información extraída del código fuente de la página principal de esta empresa, que resume perfectamente los objetivos que se plantean con su software <<http://www.metamatrix.com>>: <META NAME="description" content="MetaMatrix provides a scaleable, eCommerce platform to house and manage metadata at the enterprise level, giving developers a consolidated view of firm-wide data content. Using the central store of metadata, the MetaMatrix information integration platform allows uniform data access regardless of the native database technology.">.

los mismos. A través de lo que denominan una *plataforma de integración de metadatos*, permite homogeneizar el acceso a la información, independientemente de la tecnología que utilice la base de datos original.

- La empresa Interwoven <<http://www.interwoven.com>> que está vinculada sobre todo al desarrollo de aplicaciones relacionadas con el comercio electrónico y a la creación de portales comerciales y/o empresariales ha desarrollado, integrándolo dentro de los flujos de trabajo de su aplicación principal (TeamSite), **MetaTagger**, un módulo de software para crear metadatos. Las funcionalidades de Metattager se definen en Interwoven, de una forma más retórica, como: *automatizar el proceso de añadir inteligencia al contenido a través de la creación de metadatos ricos y descriptivos a partir de artículos, documentos o páginas Web*²⁰⁰. Según la información que aparece en la página Web de este producto, se trata de una herramienta para la clasificación automática que se basa en estándares y vocabularios controlados para crear metadatos. Aunque no hemos podido evaluar una demo de Metatagger, y a pesar de que está eminentemente dirigido al desarrollo de sitios para el comercio electrónico, quizás sea también una herramienta adaptable a la gestión de otro tipo de información no comercial como una biblioteca digital u otro sistema de información de calidad.

En cualquier caso, los orígenes de estas tecnologías están relacionados con lo que Kuhlthau²⁰¹ denomina el *paradigma bibliográfico*. Este paradigma recoge múltiples investigaciones realizadas durante décadas para desarrollar herramientas

²⁰⁰ MettaTagger <<http://www.interwoven.com/products/metatagger>>. Esta misma idea de que los metadatos añaden "inteligencia" a la búsqueda y recuperación de información digital, la vemos también en autores como Ianella. Renato Iannella. Mostly Metadata... *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA1998>

²⁰¹ C. Kuhlthau. Inside the Search Process: Information Seeking from the User's Perspective. *Journal of American Society for Information Science*, 1991, vol. 42, nº 5, p. 361-371. Citado por: Lisa D. Murphy. *Op. cit.*, p. 56.

que mejoren el acceso a la información (sea esta digital o tradicional) a través de patrones que faciliten la recuperación de información, esto es, en última instancia, metadatos. Entre todas las aplicaciones que desarrollan un módulo de metadatos como parte integral de un sistema GED o sistema de gestión de información (sea esta empresarial-comercial, o no), amerita mención especial el software de Blue Angel Technologies, **MetaStar**²⁰². Como ya hemos mencionado²⁰³, se trata una herramienta basada en XML para gestionar la información en el entorno del comercio electrónico, y para compartir información en bibliotecas digitales. Una completa *suite* de programas que integra herramientas para la creación de metadatos, la indización y la búsqueda de información. A diferencia de las dos aplicaciones anteriores, MetaStar tienen una línea comercial más directa sobre bibliotecas digitales, ya que incluye a los bibliotecarios como uno de los tipos de destinatarios a quien dirige sus productos.

Se comercializa al más puro estilo de los SIGB, de forma modular y escalable. La *suite* de programas, que en un principio (1997) estaba compuesta de tres módulos (un cliente, un servidor y un gestor de metadatos²⁰⁴) y posteriormente (1998) por siete módulos (*Data Entry, Harvester, SDKs, Search client, Gateway, Server,*

²⁰² Programa-versión: MetaStar

Homepage: <http://www.BlueAngeltech.com>

Tipo: Software integrado de gestión de metadatos (*suite*)

Precio: No consta. Realizan presupuestos ajustados al sistema de información que lo solicita.

Plataforma/requisitos:

- Microsoft Windows NT/2000 y UNIX
- Servidores Web: Apache, Microsoft Internet Information Server, y Netscape Enterprise and FastTrack Servers
- Software para gestión de bases de datos relacionales (RDBMS): Microsoft SQL Server y Oracle

Descarga: No tiene opción de descargar una demo. Hemos evaluado su funcionalidad a partir del acceso a la *Knowledge base* que facilita Blue Angel a sus clientes. *Vid. sup.* Nota 155.

²⁰³ *Cfr.* Capítulo 3, nota 21.

²⁰⁴ Margaret St. Pierre,, James Restivo. The MetaStar Suite of Metadata Tools [documento HTML]. En: *IEEE Metadata Conference (2. 1997. Maryland)*. IEEE Computer Society, 1997. Disponible en: <http://www.computer.org/proceedings/meta97/papers/mstpierre/mstpierre.html> (consultado el 18 de junio de 1998).

Repository²⁰⁵), está compuesta en la actualidad por nueve herramientas para la gestión integral de la información electrónica.

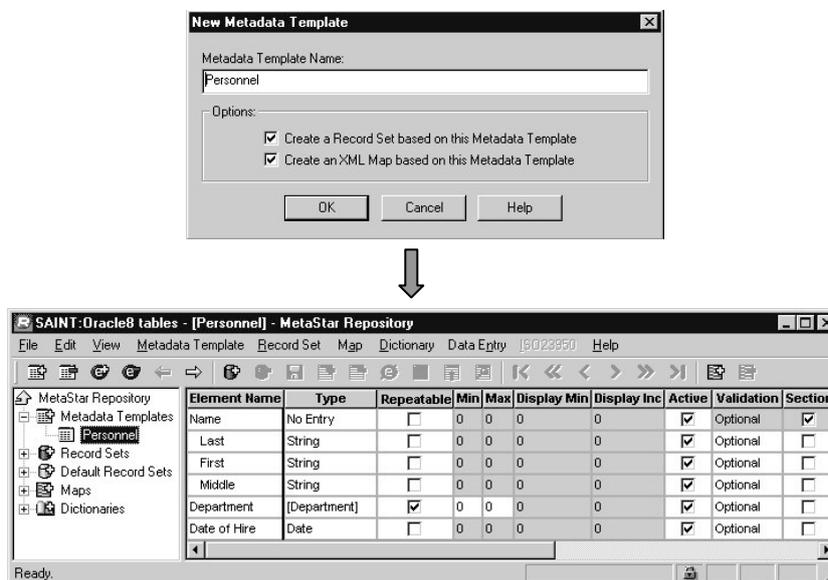


Fig. 46. Meta Star Data Entry. Creación de una plantilla de metadatos²⁰⁶

²⁰⁵ Antiguo sitio Web de la empresa: <http://www.BlueAngel.com/Products> (Versión revisada el 18 de junio de 1998 y consultado el 19 de junio de 1998).

²⁰⁶ Esta imagen procede de la *knowledge base* que incluye el soporte que da Blue Angel a los clientes <<http://www.BlueAngeltech.com/kb/Documentation/dataentry31/chapter4/section6.htm>> a la que hemos podido acceder gracias a una *password* cedida por la empresa (*Vid. sup.* Nota 155). Por ello, la plantilla se refiere a "personal" y no, a "DLO-texto" o "DLO-imagen", etc. que hubiéramos utilizado nosotros como ejemplo, ya que, aunque este software se comercialice como herramienta para la gestión del conocimiento empresarial, es perfectamente parametrizable a las necesidades de una biblioteca digital. Así se muestra a través de tres estudios de caso que presenta la propia empresa en su página Web, teniendo en cuenta las necesidades de cada biblioteca y haciendo referencia a las soluciones modulares de MetaStar y a los beneficios que reportará a: la Biblioteca Nacional de Australia <<http://www.BlueAngeltech.com/products/CaseStudies/CaseStudyNLA.htm>>, la Universidad Internacional de Florida <<http://www.BlueAngeltech.com/products/CaseStudies/CaseStudyFIU.htm>>, y al servicio TRAIL de la Comisión de archivos y bibliotecas del estado de Texas <<http://www.BlueAngeltech.com/products/CaseStudies/CaseStudyTRAIL.htm>>. Aunque también ha propuesto soluciones para servicios de información temática como el *USGS-National Spatial Data Infrastructure (NSDI) Clearinghouse* <<http://www.BlueAngeltech.com/products/CaseStudies/CaseStudyUSGS.htm>>.

A efectos de creación de metadatos, debemos destacar un módulo denominado *MetaStar Data Entry*, para seleccionar, añadir, modificar, borrar y verificar registros de metadatos (*metadata record*²⁰⁷) permitiendo que varios usuarios accedan simultáneamente, manteniendo a pesar de ello la integridad de la información en un sistema de gestión de base de datos relacional (RDBMS) (p. ej. Oracle). Una de las características más interesantes de este módulo es que permite configurar las propiedades de los registros de metainformación, los interfaces y las plantillas para crear los metadatos (Fig. 46).

Además de generar plantillas para distintos modelos de metadatos (FGDC, GILS, DIF, MARC y DC principalmente), también crea diccionarios o vocabularios controlados para mejorar la consistencia del sistema de entrada de datos. Después de que se introducen los registros y se almacenan en formato XML, entran a formar parte del *workflow* del sistema, antes de publicarse automáticamente en la WWW. Se trata pues de una herramienta muy completa y configurable a distintas necesidades de gestión de información, incluidas las bibliotecas digitales. Está implementada en Java y trabaja con los servidores Web más comunes tanto en UNIX como en Windows NT. Pero *Data Entry*, no es más que una pieza más de MetaStar, íntimamente relacionado con otros componentes de este software. Esos otros módulos son:

- *MetaStar Enterprise* que está formado por una base de datos, un motor de búsqueda y la tecnología Web que permite la indización y recuperación de información en entornos WWW. Este módulo a su vez está formado por otros cuatro subsistemas que se comercializan por separado: *MetaStar Data Entry*, del que ya hemos hablado, *Gateway*, *Repository* y *Server*.

Gateway que sirve para buscar en distintos recursos distribuidos a través de una sola sentencia de búsqueda independientemente de cuál sea el formato de estos.

²⁰⁷ Vid. Glosario.

Inclusive, Blue Angel proporciona a sus usuarios una serie de perfiles de búsqueda, según diferentes formatos (GILS, FGDC, MARC) para integrar en este módulo²⁰⁸.

Repository es un módulo para extraer y gestionar datos XML en una base de datos relacional compatible ODBC, además cuenta con un servicio que proporciona estándares preparados para importar y configurar este módulo con varias plantillas (por ejemplo MARC²⁰⁹). Además facilita la publicación en la WWW de la información extraída de los metadatos en forma de árbol jerárquico. Todo ello con una interfaz simple basada, como IE, en el paradigma SDI (*Single Document Interface*).

Server, es un completo paquete de software para buscar información en red (Internet/Intranet), para lo que utiliza un protocolo de comunicaciones normalizado ANSI/NISO Z39.50 (ISO23950), a su vez *MetaStar Server* consta

²⁰⁸ A través del soporte informativo de la empresa a sus clientes (*Vid. supr.* Nota 155), hemos accedido a los ficheros comprimidos que contienen toda la información necesaria (opciones de búsqueda según el protocolo Z39.50 —marc.xml, gils.xml, GEO22.xml, etc. y el fichero q contiene la DTD del perfil en cada caso, profile.dtd—, ficheros de configuración —config.xml, gateway.dtd—, plantillas HTML para las distintas presentaciones —brief.htm, full.htm, etc.— imágenes y las propiedades del *servlet*) para configurar distintos ejemplos de búsqueda:

- <<http://www.BlueAngeltech.com/KB/gateway/configurations/LibrarySample.zip>> (ejemplo de perfil según MARC/Z39.50).
- <<http://www.BlueAngeltech.com/KB/gateway/configurations/GILSSample.zip>> (ejemplo de perfil según GILS).
- <<http://www.BlueAngeltech.com/KB/gateway/configurations/FGDCSample36.zip>> (FGDC).
- <<http://www.BlueAngeltech.com/KB/gateway/configurations/MultiProfileSample.zip>> (Varios perfiles).

²⁰⁹ Ese servicio que proporciona Blue Angel a sus clientes incluye, entre otras prestaciones, MARC_Service_Pack

<http://www.BlueAngeltech.com/KB/Repository/Configurations/MARC_Service_Pack.zip> que contiene el fichero de configuración MARC (marc.xml) y la DTD correspondiente (metadata-template.dtd). Este pack prevé la interoperabilidad entre etiquetas MARC y XML, que será de gran utilidad si se quiere configurar el servidor de *MetaStar* para que soporte registros con distintas sintaxis (XML, MARC y GRS).

de tres componentes: un servidor, un indizador, un fichero de configuraciones. Además se integra con algunos de los motores de búsqueda más conocidos, como Altavista Search y Fulcrum SearchServer and Knowledge Network.

- Otro de los módulos principales de MetaStar es *Harvester*²¹⁰ un robot que extrae y recopila información del sitio/s que se le indiquen, ya estén estos en local o en la WWW. Extrae y analiza tanto documentos XML como HTML, recopilando información de los elementos que se definan *a priori*, por ejemplo todos o parte de los metadatos incluidos en el documento. Es un software muy flexible y configurable, tanto para la información que extrae, como para el formato de salida de los datos o metadatos indizados, ya que puede extraer información en XML y convertirla a HTML o viceversa, e incluso en simples ficheros .TXT de pares de atributo valor.
- *MetaDoor Interchage* es una herramienta original de Blue Angel que genera, por ejemplo, XML a partir de una hoja de cálculo de MS Excel, separando para ello el proceso de entrada de la estructura del registro, de tal forma que permite asignar la entrada de datos a personas no familiarizadas con XML.
- *MetaStar SDKs (Software Development Kits)* es otro módulo constituido por un conjunto de herramientas que proporciona nuevas formas de configurar y adaptar las aplicaciones. SDKs está compuesto por tres tipos de *kits* de herramientas: un cliente SDK que, sustentado sobre Z39.50 que habilita al resto de las aplicaciones realizar búsqueda y recuperación de datos estructurados y distribuidos, SDK en el servidor, que facilita la publicación de información estructurada en Internet/Intranet, y por último, *Translation SDK* que mejora las posibilidades de interpretar o escribir datos XML en las distintas aplicaciones que conforman MetaStar.

²¹⁰ Vid. Glosario, *harvest-harvesting /gather-gathering*.

- Finalmente, *Mobile* es el último módulo que ha incorporado MetaStar en su *suite*. Como su propio nombre indica crea accesos a todas las funcionalidades de MetaStar desde plataformas móviles, actualmente desde Palm Pilot VII, pero espera ser compatible con otras en el futuro.

El segundo grupo de aplicaciones que pertenecen a lo que hemos llamado SIG-DLOs está constituido por software específico, también modular, que se ha desarrollado en un marco más amplio de aplicaciones bibliotecarias, en los denominados tradicionalmente, Sistemas Integrados de Gestión de Bibliotecas. En este sentido destaca Hyperion de Sirsi y MetaLib y DigitooLibrary de Ex-Libris que surgen en ambos casos en el contexto de la gestión de recursos electrónicos en bibliotecas, sobre todo universitarias y de investigación, donde la integración de recursos digitales y convencionales es inminente²¹¹.

Sirsi fue la primera empresa en arbitrar soluciones para enfrentar el problema de la inclusión de colecciones digitales en los servicios bibliotecarios, así anunció, a principios del año 2000 su software específico para esta función. **Hyperion**²¹² está compuesto por siete módulos, dos de los cuales, *Metadata manager* y *Media text index manager*, permiten diseñar perfiles y conjuntos de propiedades para identificar y describir los recursos de la colección digital utilizando como base el modelo DC. Pero estos módulos de definición y creación de metadatos están integrados con los

²¹¹ *Vid. supr.* Nota 92.

²¹² No se trata ahora de hacer una descripción detallada de cada uno de estos software (máxime cuando no hemos podido evaluar su funcionamiento real a través de una demo). Únicamente presentamos las soluciones comerciales en el ámbito de los SIGB para la gestión de colecciones digitales, así como de demostrar que las últimas tendencias a tenor de los metadatos y de las bibliotecas digitales (2000-2001) tratan de recoger la problemática que, desde su inicio planteamos en esta tesis doctoral.

Más información sobre este software: <http://www.sirsi.com/Sirsiproducts/hyperion.html> o desde un punto de vista técnico, http://www.sirsi.com/Sirsipdfs/Hyperion_Technical_Overview.pdf Hyperion: sistema de archivos de documentos digitales [presentación PPT]. En: *Jornadas de Tecnología de la Información: Sistemas de Gestión Integrada de Bibliotecas (2000. Bahía Blanca)* [CD-ROM]. Bahía Blanca: Universidad Nacional del Sur, Biblioteca Central, 2001.

otros cinco que forman el SIG-DLOs: *Media file manager*, para la gestión de unidades de almacenamiento, de forma que la ubicación física de los documentos sea totalmente transparente para los usuarios finales; *Hierarchy manager*, crea el árbol de clasificación a partir del cual se agruparán y ordenarán los documentos; *Access manager*, gestiona las restricciones de acceso y los perfiles de seguridad para acceder a los documentos; *Import facility*, para importar grandes volúmenes de documentos; *Web client interface*, interfaz HTML a través de la cual los usuarios pueden acceder a las bases documentales gestionadas por Hyperion.

DigiTooLibrary²¹³, más reciente que Hyperion (Junio 2001), como su propio nombre indica pretende incluir todas las herramientas necesarias para solucionar el problema de construir y gestionar colecciones digitales. Se presenta como un sistema de gestión de activos digitales (*Digital Asset Management System*²¹⁴), que basa su funcionamiento, desde el punto de vista de los metadatos, en estándares, tanto de estructuras de datos (DC, MARC21, MAB, TEI, y EAD), como en estándares del contenido de los datos (TIFF, GIF, JPEG, PDF, y texto en SGML, HTML o XML) o de intercambio de información (Z39.50 o OAI) e incluso soporta la sintaxis de intercambio de RDF. Este software se integra tanto con MetaLib, como con el propio SIGB de Ex-Libris, Aleph.

MetaLib es un software para la creación de portales académicos de recursos digitales²¹⁵. Nace, como Hyperion de Sirsi en el año 2000 con la vocación, en este

²¹³ Para una descripción más completa *Vid.:* <http://www.exlibris-usa.com/digitoolibrary/index.html> y <http://www.exlibris-usa.com/uploadFiles/digitool.PDF>

²¹⁴ Esta curiosa designación no nos sorprende en el sentido de que reconoce tácitamente que el fundamento y el objeto que constituye una biblioteca digital no son sólo "documentos" digitales sino "activos digitales" lo que nosotros hemos denominado DLOs que implican también imágenes fijas o en movimiento, sonido e información textual.

²¹⁵ De una forma elemental podemos decir que MetaLib es a DigiTooLibrary lo que Metacontents Builder a TagGen (*Cfr. supr.* 4.3.2.2). Más información sobre MetaLib en: <http://www.exlibris-usa.com/metilib/overview.html> o el documento presentado en la conferencia Interfaces (México): [cont.]

caso, de aunar en un mismo sistema de información recursos digitales y tradicionales (biblioteca híbrida). Permite la organización, disseminación, recuperación y acceso a información académica en un ambiente heterogéneo de catálogos bibliográficos y bases de datos electrónicas. MetaLib puede entenderse como un módulo de Aleph, pero es en sí mismo modular. Por otra parte facilita la ampliación a otros servicios extendidos a través de la tecnología propietaria de Ex-Libris SXF, que genera enlaces dinámicos a través de un análisis de los documentos recuperados y del entorno de consulta de los usuarios, y aumentar de esta forma el número de ítems recuperados.



Fig. 47. Esquema genérico de los módulos de un SIG-DLOS²¹⁶

4.3.2.4. Otras aplicaciones de creación de metadatos

Aunque parece que ya hemos planteado una visión de todas las posibilidades de herramientas informáticas para la creación de metadatos, debemos señalar otras específicas que, si bien podrían integrarse en alguna de las categorías que hemos

Alfredo Bronsoiler Frid. De sistemas de bibliotecas a sistemas de información [documento PDF]. En: *Interfaces 2000: Foro internacional sobre biblioteca digital (2000. Colima)*. Disponible en: <http://www.ucol.mx/interfaces/interfaces2000/mesast/Mt01.pdf> (consultado el 8 de agosto de 2001).

²¹⁶ La imagen anterior resume, a través un esquema genérico, la composición de estos SIG-DLOS cuando surgen en entornos más consolidados de software como son los SIGB.

señalado en los dos apartados anteriores en virtud a su funcionamiento —plantillas *online* (4.3.2.1.), software cliente (4.3.2.2.), o parte modular de un SIG-DLOs (4.3.2.3)— por su particularidad, bien del contexto temático en el que surgen, de su utilidad para transformar datos o de la sintaxis de representación que utilizan, deben considerarse aparte. Por ello, dentro de este marbete tan amplio de "otras aplicaciones" trataremos de resumir otras de las herramientas relacionadas con la creación de metadatos atendiendo a dos grupos fundamentales:

- a) Aplicaciones de un dominio temático específico.
- b) Conversores de metadatos.

Además de estos dos grupos fundamentales también podríamos incluir dentro de "otras herramientas" aquellas vinculadas a la especificidad de la sintaxis y al modelo RDF, marco en el que han surgido múltiples validadores, analizadores y herramientas de edición específica. Sin embargo hablaremos de ellas al tratar el *Resource Description Framework* como un metamodelo de metainformación, en el apartado 5.2.3.

a) Aplicaciones de un dominio temático específico

En casi todas las herramientas genéricas para la creación de metadatos que hemos analizado vemos un sesgo importante, atendiendo a los formatos que soportan. Tanto las plantillas, como el software cliente, están dirigidas a la creación de metadatos compatibles con el estándar Dublin Core o formatos basados en éste, como AGLS, en el caso del software australiano, o como mucho, en el mejor de los casos, sobre todo en módulos de creación de metadatos en SIG-DLOs están volcados a la compatibilización de varios formatos. Vamos a mencionar ahora algunas aplicaciones que surgen con la vocación de soportar un esquema de metadatos propósito específico.

Como ya hemos dicho, el software relacionado con la creación de metadatos, a pesar del lento desarrollo de aplicaciones comerciales, se ha dado especialmente en

entornos de investigación y normalmente vinculado a proyectos de organización de información particulares. Por ello, a pesar de que al hablar de las plantillas ya mencionamos algunas de ellas para la creación de un tipo de metadatos específico, como el *Medical Metadata Creator* <<http://medir.ohsu.edu/bicc-informatics/ebm/latest.htm>>, existen otras muchas aplicaciones que sirven para crear también metadatos específicos pero dentro de un proyecto o modelo concreto, bien como un módulo dentro de lo que hemos denominado sistema integrado de gestión de información digital, bien como software cliente o *template* en línea e individual (*stand-alone*). Nos referimos a todas aquellas herramientas de propósito específico²¹⁷ desarrolladas *ad hoc* para ayudar a la asignación de metadatos según un esquema particular o particularizado y en un contexto informativo temático.

Dentro de esta categoría debemos destacar aquellas vinculadas a la información geoespacial, que es el ámbito informativo que más software específico ha desarrollado. Desde aplicaciones para editar y registrar metadatos, hasta herramientas para el pre-post/procesamiento o validación de los mismos²¹⁸. La mayor parte de estos software se han diseñado para ayudar a la creación de metadatos conformes al

²¹⁷ "Metadatos de propósito específico" es la denominación genérica que hemos elegido para englobar aquellos esquemas o modelos desarrollados para la descripción de una información temática (Cfr. 5.2.2.). Por lo mismo, y en consonancia con este tipo de formatos de metainformación que analizaremos en el capítulo siguiente, podemos también hablar aquí de aplicaciones de propósito específico.

²¹⁸ La homogeneidad que se persigue desde hace mucho tiempo en los Sistemas de Información Geográfica explica que los listados de herramientas para la gestión de metadatos geoespaciales sean muy parecidos y recogen los software más usados para el tratamiento de metadatos geoespaciales.

Una recopilación, exhaustiva y actualizada, que describe las aplicaciones más importantes en este ámbito es la realizada por Hugh Phillips en el marco del *clearinghouse* de información geoespacial de Wisconsin (WISCLIC). Hugh Phillips. *Metadata Tools for Geospatial Data* [documento HTML]. Wisconsin Land Information Clearinghouse, rev. 17 de julio de 2001. Disponible en: <http://wisclinc.state.wi.us/metadata/mtools.html> y en: <http://badger.state.wi.us/agencies/wlib/sco/metatool/mtools.htm> (consultado el 30 de julio de 2001). La versión en español que recoge estas herramientas se encuentra ubicada en el homónimo uruguayo del WISCLIC, en: http://www.clearinghouse.com.uy/ppoint/metatool/mtools_es.html

estándar CSDGM (*Content Standards for Digital Geospatial Metadata*) del FGDC aunque algunos de ellos se ajustan a perfiles de metadatos de un sistema de información particular. Este es el caso por ejemplo, de **Metalite**²¹⁹, una de las aplicaciones más conocidas y utilizadas en este ámbito, destinada a recopilar y validar metadatos adaptados al estándar del *Federal Geographic Data Committee* (FGDC). Se desarrolló en el marco concreto del *United States Geological Survey* (USGS) y del Programa de Medioambiente de las Naciones Unidas (UNEP) para documentar datos espaciales para el uso en sistemas de información geográfica, de tal forma que se mejore el intercambio de información espacial de interés para la gestión de recursos geológicos y medioambientales.

Metalite es un software independiente (*stand-alone*) que opera en una base de datos Access. Su sistema de entrada de datos consta de una interfaz muy sencilla y multilingüe que permite trabajar con un subconjunto de elementos del FGDC²²⁰ y utiliza un analizador (MP: *Metadata Parser*) para validar los metadatos que genera y permite exportar los registros en SGML, XML, HTML, o TXT.

²¹⁹ Programa-versión: MetaLite 1.7.5

Homepage: <http://edcnts11.cr.usgs.gov/metalite>

Tipo: Freeware

Plataforma/requisitos: Win 95/98 y NT

Nombre (tamaño) del fichero: meta175.exe (4606 Kb)

Actualización del 27/07/01: metaliteutilities.zip (54 Kb9)

Descarga: <http://edcnts11.cr.usgs.gov/metalite/downloads/meta175.exe>

<http://edcnts11.cr.usgs.gov/metalite/downloads/metaliteutilities.zip> (actualización)

Fecha de descarga: 22/03/01 18:06; 07/08/01 19:01 (actualización)

²²⁰ La Fig 48, a continuación, muestra por ejemplo, la introducción de las coordenadas geográficas que contempla el *schema* para la descripción de este tipo de información (FGDC):

```
Bounding_Coordinates= West_Bounding_Coordinate +;
```

```
East_Bounding_Coordinate +; North_Bounding_Coordinate +;
```

```
South_Bounding_Coordinate
```

Una de las particularidades más curiosas de este software es el hecho de que permite trabajar a un nivel multilingüe (inglés, francés, español y portugués), tanto en la interfaz en general, como en los descriptores de materia que contiene, lo que corrobora la internacionalización de la información geoespacial en general, y de este software

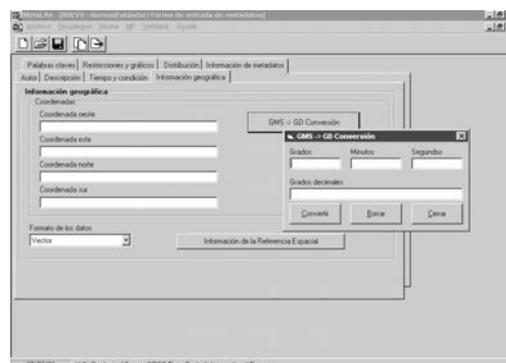


Fig. 48. Metalite (español)

en particular. Metalite además ha servido de inspiración para el desarrollo de otras aplicaciones vinculadas a proyectos particulares de información geográfica o espacial, como es el caso de NJMetalite²²¹ que se ha creado según las necesidades concretas de la información del GIS (*Geographical Information System*) de New Jersey.

A este mismo nivel de aplicaciones específicas podemos destacar también otros software desarrollados para otros formatos de metainformación o para otros entornos temáticos. Vamos a poner a continuación algunos ejemplos representativos de toda la casuística que nos podemos encontrar en esta categoría.

Mathematics Metadata Mark-up Editor²²² (MMM 3.0) <<http://www.mathematik.uni-osnabrueck.de/cgi-bin/MMM3.0.cgi>> por su parte, está dirigido a un tipo y temática de información particular: artículos y *preprints* científicos sobre matemáticas. Es una sencilla plantilla de JavaScript desarrollada en la universidad alemana de Osnabrueck que facilita la creación de registros de

²²¹ Para mayor información y descarga de este software, Vid. <http://www.state.nj.us/dep/gis/endex/njml.htm>

²²² Esta aplicación para la creación de metadatos asociados a DLOs sobre matemáticas, también puede descargarse en: <ftp://ftp.mathematik.uni-osnabrueck.de/pub/MMM> y posteriormente, instalarse en un servidor Web propio en el directorio cgi-bin.

metadatos para ese tipo de recursos informativos de contenido matemático. El hecho de estar vinculado a un campo temático tan concreto permite asignar las materias según sistemas de clasificación específicos: el MSC (*Mathematics Subject Classification*) que está integrada dentro de la aplicación pudiendo elegir las materias primaria y secundaria, o a través de la consulta de otros *schemes* como el ZDM (*International Reviews on Mathematical Education*), el *Computing Classification System*, o PACS. Una vez completada la información (al menos título, autor y clasificación) edita los metadatos en forma de página Web cuyo código fuente es un completísimo registro en RDF-XML según el esquema DC y sus calificadores.

En el ámbito de la información sobre Educación destacan el conjunto de módulos desarrollados en el contexto del proyecto GEM²²³ (*Gateway for Education Materials*). Uno de esos módulos es **GEMCat**²²⁴, denominado módulo de catalogación como en los SIGB convencionales. Desarrollado en la Universidad de Siracusa, facilita la creación de metadatos según el esquema de veintitrés elementos del GEM.

²²³ El GEM es un consorcio que ha desarrollado una serie de estándares (entre otros, un modelo de metadatos propio que consta de veintitrés elementos, vocabularios controlados, etc.) y de mecanismos técnicos para crear un sistema eficaz de acceso a los materiales educativos de la Red. Las herramientas que han desarrollado para llevar a cabo ese proyecto, además del módulo de catalogación que mencionamos aquí (GEMCat), son: un software que indiza y recoge los metadatos generados por GEMCat (GEM Harvest), y un sistema que genera automáticamente páginas HTML para esos metadatos (GEMBrowseBuilder).

²²⁴ Programa-versión: GEMCat 3.22

Homepage: http://www.geminfo.org/Workbench/Workbench_cataloging.html

Tipo: Freeware

Plataforma/requisitos: Win9x/NT/2000, Mac OS 7.6.1+, Solaris y Un*x. Requiere *Java Runtime Environment* (JRE) ver. 1.1.5 o superior

Nombre (tamaño) del fichero: install.exe (3981 Kb)

Descarga: <http://128.230.185.59/software/gemcat/InstData/Windows/VM/install.exe>

Fecha de descarga: 08/08/01 04:34

Acceso Online (plantilla): <http://128.230.185.59/software/gemcat/index.html>

La introducción del contenido de los distintos elementos está muy controlada a través de varios schemes que aparecen implícitos en la aplicación y que coadyuvan a crear un sistema sólido, coherente y eficaz de metainformación. Uno de estos estándares de contenido es el Tesoro de educación ERIC, que se consulta como un submódulo dentro de GEMCat (la Fig. 49 muestra la interfaz de creación de metadatos y en la parte inferior derecha aparece la herramienta del tesaurero). A la hora de salvar la información, ésta se puede guardar junto con el fichero HTML al que describe o de forma separada haciendo así funciones de editor o de generador, según la diferenciación que comentábamos al comienzo del apartado 4.3.2. Además, el hecho de que GEMCat sea un módulo del proyecto también nos podría llevar a clasificarlo dentro de las soluciones integradas, ya que por ejemplo, los metadatos que crea serán indizados por otro de los módulos, GEMHarvester. Sin embargo hemos decidido tratarlo aquí pues su característica principal es que se trata de un software ad hoc para asignar metadatos a la información educativa, según un formato particular.

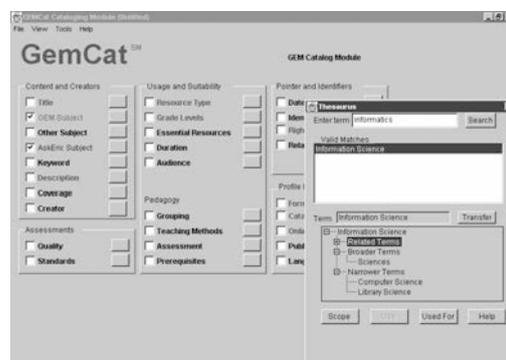


Fig. 49. GEMCat (Thesaurus)

Finalmente, dentro de esta categoría de "otro software" que sigue un esquema de metainformación específico, debemos mencionar **MARCit**²²⁵, que a pesar de las críticas que se pueden hacer al MARC para catalogar Internet, utiliza este formato

²²⁵ Programa-versión: MARCIt

Homepage: <http://www.marcit.com>

Tipo: Demo

Precio: \$49.95

Plataforma/requisitos: Win 9x o NT

Nombre (tamaño) del fichero: MARCItSampleSetup.exe (1966 Kb)

Descarga: <http://www.marcit.com/download/MARCItSampleSetup.exe>

Fecha de descarga: 25/04/01 21:59

como esquema de metadatos. A diferencia de las anteriores, es una aplicación comercial de la empresa Sagebrush <<http://www.sagebrushcorp.com>> dedicada a la automatización de bibliotecas escolares.

Su funcionamiento es muy sencillo: una vez visualizado en el navegador el recurso que se quiere catalogar, se ejecuta MARCIt que incluye lo que denominan un icono  flotante que al pulsarlo traspasa al programa, bien automáticamente bien porque se ha configurado previamente (Fig. 50), los datos sobre el DLO que se quiere

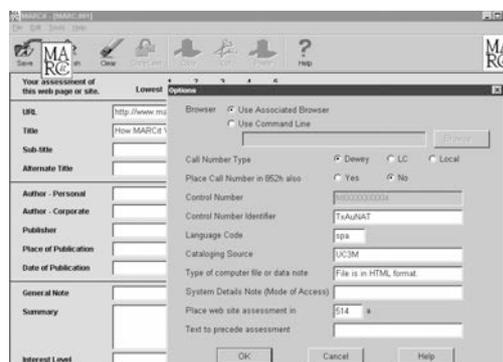


Fig. 50. MARCIt (opciones de configuración)

catalogar. Éstos serán los datos básicos de la catalogación en MARC: cabecera (LDR), el número de control (001) y el identificador de éste (003), la fecha y hora de la última transacción (005), el campo fijo de descripción física (007), el 008, la fuente de la catalogación (040), el título (245), la designación del tipo de material (245\$h), las características del fichero (256\$a), y varias notas 514\$a (calidad de los datos), tipo (516\$a), forma de acceso (538\$a), y el campo MARC específico de los recursos electrónicos donde se consigna el URL (856\$u), a los que se añadirá la información relativa al resto de los elementos. Finalmente se salva el registro MARC que podrá importarse al sistema de automatización de bibliotecas y de esa forma integrar en el catálogo los DLOs con el resto de documentos incluidos en la base de datos bibliográfica que, como vimos en la primera parte de este capítulo (4.1), es la principal ventaja de usar MARC como formato de codificación de metadatos.

b) Conversores de metadatos.

El hecho de que el MARC sea el formato por excelencia en todas las bibliotecas y servicios de información tradicionales hace que existan sistemas de catalogación de documentos electrónicos como MARCIt, que trabaja originalmente con MARC como modelo de metadatos, pero también aplicaciones que trabajan a

priori con otros esquemas pero que incluyen el MARC como un formato de conversión de metadatos²²⁶. El nivel de consolidación del estándar MARC como formato de intercambio bibliográfico hace que existan también herramientas que sirven específicamente para convertir de otros esquemas de metainformación a dicho formato bibliotecario o de otros modelos de metainformación a MARC. Son los que denominamos, de forma genérica, conversores de metadatos²²⁷. Estos sistemas de conversión serán esenciales cuando el OPAC de una biblioteca se utilice para acceder a depósitos de metadatos almacenados en bases de datos locales, sobre todo por tres razones que señala Amanda Xu²²⁸:

- *Para recoger, actualizar y convertir metadatos en el OPAC.*
- *Para integrar en el OPAC los metadatos creados localmente.*
- *Para visualizar los metadatos en un formato específico [y uniforme] para el usuario.*

La integración de los datos descriptivos de los recursos electrónicos en catálogos MARC convencionales es una de las principales preocupaciones en muchos contextos bibliotecarios. Esto ha hecho que se desarrollen plantillas y software para automatizar la conversión de registros de metadatos a registros MARC. Una de las herramientas más conocidas en este sentido es el *Dublin Core to MARC Converter (d2m)* <<http://www.bibsys.no/meta/d2m>>, creado en el marco del *Nordic*

²²⁶ Por ejemplo, herramientas como TagGen Profesional, contemplan esta posibilidad (*Cfr. sup.* Fig. 37.). Los sistemas integrados de gestión de la información digital como MetaStar, también tienen en cuenta la compatibilidad MARC de los registros de metadatos creados originalmente en XML (*Cfr. sup.* Nota 209). Asimismo el proyecto CORC de OCLC permite generar metadatos asociados a DLOs tanto en formato DC como en MARC y pasar de una a otra visualización gracias a las tecnologías XML implícitas.

²²⁷ *Cfr. sup.* Nota 32.

²²⁸ Amanda Xu. *Metadata Conversion and the Library OPAC* [documento HTML]. Massachusetts: Institute of Technology, rev. 30 de enero de 1998; The Haworth Press, 1997. Disponible en: <http://web.mit.edu/waynej/www/xu.html> (consultado el 25 de junio de 2001).

Metadata Project por BIBSYS (un sistema compartido de información para todas las bibliotecas universitarias noruegas) la Biblioteca Nacional de este país y varias bibliotecas de investigación. Esta aplicación que funciona *online* es capaz de convertir metadatos DC en sintaxis HTML (metaetiquetas) a registros en formato USMARC y en todos los formatos MARC específicos de los países nórdicos.



Fig. 51. d2m (Dublin Core to MARC converter)



Fig. 52. d2m. Resultados en formato TXT

Las figuras 51 y 52, muestran la conversión del código HTML de los metadatos relativos a esta tesis doctoral como si fuese un DLO para la WWW, generados anteriormente con Reggie, al formato USMARC. Los datos en MARC se pueden presentar en TXT (Fig. 52), HTML o bien convertirlos directamente al archivo binario metadata de la ISO 2709.

Los conversores de metadatos son cada vez más frecuentes y persiguen la interoperabilidad entre distintos formatos de metainformación. Dada la cantidad de esquemas, estándares y modelos de metadatos que están surgiendo en distintos dominios informativos, la interoperabilidad entre todos ellos y entre cada uno de los esquemas y el MARC será uno de los principales retos de la investigación en el ámbito de los metadatos²²⁹.

²²⁹ Vid. Capítulos 5-6 y 8.3.2.

A pesar de que los conversores más comunes se crean para cambiar de DC a MARC o viceversa, también hay otros que, por su especificidad en los formatos con los que trabajan, podríamos haberlos clasificado en el grupo anterior (aplicaciones de un dominio temático específico). Tal es el caso, por ejemplo de **VT-2-FED Metadata Converter** que sirve para trasladar al formato estándar del FGDC metadatos específicos del GIS de Vermont²³⁰.

4.3.3. Consideraciones para la elección de un software

Para concluir, podemos decir que los metadatos deberían asignarse con alguna de las herramientas que hemos comentado a lo largo de todo este apartado 4.3. La elección del software de creación de metadatos dependerá principalmente de: a) si somos los creadores de un recurso Web y necesitamos una herramienta que nos facilite las metaetiquetas mínimas que entienden los buscadores generales, simplemente para aumentar su visibilidad en estos sistemas de búsqueda, y con ello, aumentar el tráfico en nuestros sitios Web; o b) si lo que queremos es crear metainformación según un estándar (*schema*) para construir una biblioteca digital perfectamente organizada o un servicio de información de calidad, en cuyo caso el software de edición y creación de metadatos será sólo una parte o un módulo de un sistema informático de almacenamiento, búsqueda y gestión de metadatos más complejo. En el segundo caso (b), tendremos que tener en cuenta también 1) si somos los creadores de la información digital o 2) los gestores documentales de la misma²³¹, o 3) un agente externo al proceso de producción del DLO²³². De todas formas, a tenor de las herramientas que hemos analizado, y que representan una muestra de la

²³⁰ Más información sobre el funcionamiento y conversión de metadatos en este Sistema de Información Geográfica de Vermont, en: <http://geo-vt.uvm.edu/cfdev2/Library/meta/ccap.cfm>

²³¹ Los *repository managers* que dirían Dempsey y Heery. *Cfr. supr.* Nota 55.

²³² Esta diferenciación marca un criterio para tipificar las bibliotecas digitales sobre el que abundaremos en el capítulo 8 y también en el proyecto de aplicación a la BVC del capítulo 9.

realidad del mercado del software, podemos decir que las características que debe cumplir un software para la creación de metadatos deben ser:

- En cualquier caso:
 - Que tenga una interfaz amigable y sencilla.
 - Que permita la consulta de un manual de ayuda con información sobre el estándar/es que utiliza, las etiquetas que soporta y su valor según un esquema determinado, así como información sobre el manejo de la aplicación en sí misma.
 - Que permitan visualizar la página para la cual se asigna las metaetiquetas o los metadatos, bien a través de un navegador propio, o a través de la integración con AU instalados (*v. gr.* IE).
 - Que permitan ver el código fuente a través de un editor de textos.
 - Que permita repetir etiquetas.
 - Que facilite la inclusión de metaetiquetas *ad hoc*.
- En el caso de que su finalidad sea aumentar las posibilidades de recuperación en los motores de búsqueda genéricos (a), el software debe permitir:
 - Realizar un control del contenido de las metaetiquetas (control de *spamming*, densidad de los descriptores, etc.).
 - La validación de palabras clave en el caso de que éstas se extraigan automáticamente.
 - El envío de la página para la que se han creado las etiquetas <META> a distintos motores de búsqueda simultáneamente.

- En el caso de que su finalidad sea facilitar la edición de metadatos concretos, según un modelo particular, o varios de ellos, que soporte la estructuración de la información en una biblioteca digital (b), debemos considerar:
 - Si trabaja con plantillas, fórmula muy común de este tipo de aplicaciones, que éstas permitan incluir aspectos de sintaxis (elementos, subelementos, calificadores, etc.).
 - Que permitan la validación de los elementos obligatorios del modelo o de los seleccionados *ad hoc* para una biblioteca digital particular.
 - Que facilite cierto nivel de control tanto en las autoridades, distinguiendo entre autores corporativos y de carácter personal, como en las materias, permitiendo seleccionar un esquema de vocabulario controlado o tesaurus y un esquema de clasificación (*schemes*), e incluso varios de ellos simultáneamente.
 - Que permitan trabajar, esto es crear y/o editar, tanto esquemas *schemas* de metadatos, como *schemes* para controlar el contenido de las metaetiquetas.
 - Que faciliten la elección de etiquetas o del contenido de éstas de una lista de opciones. Esto es que permita la integración de varios *schemas* y *schemes*.
 - Que sea lo suficientemente flexible para hacer frente a los cambios que sufran los distintos estándares de metadatos que, en la mayoría de los casos —como hemos anticipado y que trataremos profusamente en el capítulo siguiente— están todavía en fase de desarrollo²³³.

²³³ Según Waugh, el problema crítico en las aplicaciones de metadatos es justamente la flexibilidad. Este autor denuncia —y se puede corroborar en el análisis de herramientas que hemos presentado [cont.]

- Que permita varios formatos de codificación de metadatos y distintas sintaxis de visualización (HTML, XML/RDF, inclusive también XHTML).
- Que pueda integrarse fácilmente, tanto dentro de un sistema de almacenamiento de metainformación, como de búsqueda y recuperación de metadatos.

Hasta aquí hemos analizado las herramientas informáticas vinculadas al proceso de creación de metadatos. Aunque mayoritariamente hemos evaluando el software desde el punto de vista de la creación o asignación de metadatos, debemos matizar que, salvo para meros procesos de edición de páginas Web, en los cuales una herramienta de edición de códigos fuente de metadatos, como las que hemos analizado en el apartado 4.3.1., sería suficiente, es preciso contar con sistemas de software más complejos como MetaStar o como las soluciones integradas que plantean los SIGB en los módulos para la gestión de DLOs.

Esta reflexión en torno a las aplicaciones de metadatos sirve para corroborar la distinción que sustentamos en toda esta tesis: de un lado, los metadatos dirigidos a sistemas de recuperación de información general en cuyo caso el único compromiso (incontrolable para los gestores de información digital) recae en los creadores de

aquí— que, en los actuales software de metadatos, el estándar está embebido en el propio código, con lo cual, hacer cambios es difícil y caro; por ello, el componente clave para soportar aplicaciones de metadatos flexibles es el software que pueda configurarse dinámicamente a través de la especificación previa del estándar de metadatos. Andrew Waugh. Specifying Metadata Standards for Metadata Tool Configuration [documento HTML]. En: *International World Wide Web Conference (7. 1998. Brisbane)*, rev. 9 de septiembre de 1998. Disponible en: <http://www7.scu.edu.au/programme/fullpapers/1913/com1913.htm> (consultado el 18 de marzo de 2001).

Lo más próximo a este modelo que propone Waugh, es Reggie Metadata Editor, que como vimos (Fig. 25), permite seleccionar previamente un esquema de metainformación, readaptándose la aplicación a dicho estándar. El problema que trata Andrew Waugh y que hemos reflejado también aquí, no es otra cosa que el eterno temor bibliotecario a las reconversiones del catálogo, tanto por la complejidad como por el coste que entrañan. En el caso de los metadatos, este temor aumenta al multiplicarse los formatos de metainformación y su forma de aplicación, así como la necesidad de incluir varios de ellos en un mismo sistema de información. Retomaremos también este tema al plantear la aplicación de metadatos en la BVC. *Vid.* Capítulo 9.

páginas Web para que incluyan metainformación; y de otro, el papel de los metadatos como núcleo estructural de un sistema de información de calidad o biblioteca digital, donde los metadatos deben crearse de manera normalizada, integrada y en coherencia con otras aplicaciones como los sistemas de indización (*harvesting*²³⁴) y almacenamiento que posibilitarán la recuperación precisa, coherente y relevante.

²³⁴ Vid. Glosario, *harvest-harvesting /gather-gathering*.

PARTE II:

Normalización y metadatos: formatos, esquemas o estándares

5. Modelos de metadatos. 5.1. Lenguajes de marcas y estructuración de la información. 5.2. Propuestas e iniciativas de Metadatos. **6. Normalización.** 6.1. Metadatos y el problema de la estandarización. 6.2. Estándares de metadatos: casuística y formalización 6.3. Desarrollo de un estándar formal: el caso del DCMES. 6.4. Aproximación a la normalización del vocabulario.

CAPÍTULO 5:

MODELOS DE METADATOS

Aunque en los dos capítulos precedentes ya hemos mencionado varios modelos de metadatos, plantaremos ahora una aproximación más profunda, descriptiva y crítica sobre los principales esquemas a partir de la realidad de los lenguajes sobre los que se constituye la World Wide Web. Para ello, partimos del concepto de general de marcado de documentos y de sus implicaciones en la estructuración de la información Web. En primer lugar, realizaremos una reflexión en torno a la estructuración de la información a través del etiquetado o marcado de documentos. Para ello, hablaremos del *Standard Generalized Markup Language* (SGML), como metalenguaje y marco normativo básico que permite señalar la estructura lógica de los documentos a través del marcado o balizado¹ del texto; además, trataremos la transición de la creación de contenidos en la WWW a XML (*eXtensible Markup Language*), que también es un metalenguaje de descripción de documentos, sucesor de HTML en la creación de la Web Semántica². Como colofón de la primera parte de este capítulo, dedicaremos un apartado al intercambio de información de recursos electrónicos ampliamente estructurados, que aportará un contexto básico en la justificación del marcado de documentos.

¹ Sobre el término *Mark-up/Markup language*, Vid. Glosario.

² La Web semántica es un proyecto impulsado por el propio inventor de la Web Tim Berners-Lee, y James Hendler, profesor de la Universidad de Maryland e investigador de DARPA (*Defense Advanced Research Projects Agency*) e implica la creación de una Web de datos con significado, en el sentido de que un programa de ordenador pueda aprender tanto acerca de lo que quieren decir los datos, como acerca de la información necesaria para procesarlos.

En la segunda parte del capítulo, trazaremos una síntesis descriptiva y crítica de algunos de los principales modelos de metadatos que elegimos como paradigmáticos por ser los más representativos y utilizados dentro de la amplia cantidad de esquemas que se están desarrollando en este ámbito de la descripción de la información de la Red. Distinguiremos modelos de propósito general, como el *Dublin Core Metadata Initiative* (DCMI o DC), y modelos de propósito específico, como la cabecera TEI (*Text Encoding Initiative*). De forma individual trataremos el *Resource Description Framework* (RDF) como un "metamodelo" de metadatos o un marco sintáctico para el almacenamiento e interoperabilidad de metainformación. Se pretende dar cuenta de la materialización del concepto de metadatos en propuestas concretas, así como revelar en qué medida estos formatos y estándares recogen las necesidades de información de los usuarios ante la Red.

Los motivos que han llevado a seleccionar y tratar dichos modelos de metadatos son de diversa índole: por ser los más conocidos a nivel general, como las metaetiquetas en la sintaxis HTML con valores *keywords* y *description* que interpretan algunos buscadores; por su amplia aceptación y por ser el mentor tácito de la Web Semántica, en el caso del Dublin Core; o por la representatividad que tiene para los objetivos de esta tesis, como es el caso de TEI Header³; y todos ellos, por ser formatos que responden, desde distintos niveles y ámbitos, a la complejidad y variedad de la información susceptible de transmisión digital. Mencionaremos, además, otros esquemas específicos de un tipo de información o dirigidos a una Comunidad Virtual de Usuarios o Red Temática⁴, lo que no obsta tampoco para que

³ El hecho de analizar la TEI Header se debe a que la biblioteca digital que hemos elegido para realizar un proyecto de aplicación de metadatos a tenor de esta tesis (capítulo 9) es la Biblioteca Virtual *Miguel de Cervantes*, cuya colección está formada fundamentalmente por obras de carácter humanístico, y en especial creaciones literarias. Utiliza además una reducción de la DTD de TEI para el proceso de marcado de edición digital de los documentos que forman su colección. *Cfr.* 9.1.2.

⁴ Estos colectivos son los mismos que en el sistema de creación automática de metadatos Klarity se denominan CoI (*Comunities of Interest*). *Cfr.* 4.2. Se refieren al conjunto de usuarios con intereses informativos comunes, normalmente vinculados a una disciplina científica particular.

existan otros muchos⁵ con distinto propósito o abocados a otros campos de aplicación, que no tratamos ahora para no exceder los límites de este trabajo de investigación. Por otra parte, es preciso matizar que los metadatos son tan específicos de una disciplina como su vocabulario, por eso existen casi tantos modelos como disciplinas que pretenden dar servicios de información a través de la Web.

Por último, tenemos que señalar también una salvedad: a pesar de que, en el capítulo anterior, hemos hablado del formato MARC aplicado a los primeros proyectos de organización de la Red e incluso entendido como un modelo de metadatos más, no lo tratamos así ahora, si bien son muchos los autores⁶ que se refieren al MARC como tal, e incluso como el primer formato de metainformación. Analizaremos exclusivamente los esquemas y soluciones que han surgido a raíz de la realidad de Internet, donde las distintas comunidades científicas se han preocupado por desarrollar sus propios esquemas con unos fines muy concretos: organizar, intercambiar y recuperar su información. Es cierto, no obstante, que la tradición del MARC en el ámbito bibliotecario y de la Documentación hacen de él un referente obligado a la hora de hablar de estructuración de la información, y en este sentido

⁵ Un estudio comparativo y profundo, aunque desactualizado desde nuestro punto de vista, sobre los distintos modelos de metadatos existentes, que puede completar esta aproximación a formatos concretos de metainformación en el mundo Web, es el de Dempsey y Heery que ya hemos citado en otras partes de esta investigación, por su importancia en este tema. Lorcan Dempsey and Rachel Heery. *A Review of Metadata...* *Op. cit.*, <http://www.ukoln.ac.uk/metadata/desire/overview/overview.rtf>

Si bien el documento de Dempsey y Heery es exhaustivo, su desactualización se debe a que, a pesar de que cabía esperar una revisión (pues está publicado con un número de versión —1.0—), se trata de un estudio realizado en marzo del 1997 dentro de la primera fase del proyecto DESIRE. Aunque la fecha de actualización de la página que contiene los enlaces a las versiones HTML, RTF y PDF del documento es el 28 de enero de 1999, los ficheros de contenido no han sufrido ninguna modificación ni actualización desde el 10 de junio de 1998. Por ello, en la tabla 8 basada, entre otras publicaciones de dichos autores, en esta revisión de modelos de metadatos, introducimos algunos formatos que se introdujeron con posterioridad. *Cfr.* Capítulo 3, Tabla 8.

⁶ Dovey reconoce que para los autores que pertenecen a la *escuela de la catalogación* (*Cfr.* 3.2.), el formato MARC es un estándar perfectamente viable para los metadatos asociados a objetos digitales. Matthew J. Dovey. "Stuff"... *Op. cit.*, p. 7.

aparecerá como formato de referencia, tanto en este capítulo como en el resto de este trabajo de investigación.

5.1. Lenguajes de marcas y estructuración de la información

Hasta hace no mucho tiempo, por información estructurada se entendía el conjunto de recursos que podían recuperarse a través de lenguajes de interrogación perfectamente definidos, como los de las bases de datos relacionales o las bases de datos orientadas a objetos. Por otra parte, la información semi-estructurada hacía referencia a los recursos que contenían la estructura suficiente para tratarse como bases de datos, aunque no tuvieran la potencia de un lenguaje de interrogación (como SQL por ejemplo). En el contexto de la WWW, esta estructura se establece, bien a través de una interfaz de interrogación basada en un formulario, bien a través del marcado HTML o de otro marcado textual.

Antes de hablar de la estructuración de la información en el contexto del estándar SGML (Lenguaje Genérico de Marcas Estandarizado), o del nuevo metalenguaje de marcado XML, es necesario aclarar el concepto clave de lenguaje de marcas (*markup language*⁷), que protagoniza la nueva concepción de estructuración de la información en detrimento de los lenguajes de formato y de los lenguajes de programación⁸. Ante el reto de la organización y recuperación de información electrónica, los lenguajes de marcas se han convertido en un nuevo alfabeto semántico para los productores y para los gestores de información. El lenguaje de marcas es *el sistema formal por el cual se añade información o codificación —una notación— a la forma electrónica de un documento, de tal manera que represente su*

⁷ Cfr. Glosario, *mark-up/markup language*.

⁸ La concepción de lenguaje de marca por oposición a los lenguajes de formato y a los de programación, la encontramos por ejemplo, en uno de los primeros manuales en español sobre XML. Natanya Pitts. *XML*. Madrid: Anaya Multimedia, 1999, p. 49-50, donde Pitts define sencillamente el uso de la marca *para describir las partes de un documento*.

significado y permita, así, controlar su procesamiento (Burnard y Light⁹). Lou Burnard¹⁰, uno de los padres de TEI y una autoridad científica de SGML/XML, define también el marcado de documentos como *hacer explícita una interpretación del texto*. Tradicionalmente se utilizaba para describir las anotaciones u otras marcas en un texto de tal forma que ayudaran a determinar como debía presentarse o imprimirse. De forma análoga, siguiendo a Peis y Moya¹¹, podemos identificar un *markup language* con un conjunto de convenciones de descripción de documentos (*markup*) que se usan juntas para codificar textos. Los lenguajes de esta naturaleza deben especificar: las marcas admitidas, las marcas indispensables, cómo se distinguen éstas del texto y qué significa cada una de ellas.

Normalmente —por ejemplo Gaynor¹² o Maler y El Andaloussi¹³— al hablar del marcado, etiquetado o balizado (*markup*) de la información electrónica, se distingue entre:

⁹ Lou Burnard and Richard Light. Three SGML Metadata Formats: TEI, EAD, and CIMI [documento HTML]. En: *Work Package 1 of Telematics for Libraries Project BIBLINK (LB4034)*. Bath: UKOLN Metadata Group, diciembre de 1996, rev. 14 de mayo de 1998. Disponible en: <http://www.ukoln.ac.uk/metadata/BIBLINK/wp1/sgml> (consultado el 24 de junio de 1998). Estos autores se centran en un tipo particular de documentos electrónicos estructurados: la descripción bibliográfica detallada (que es la perspectiva del SGML que nos interesa en el contexto de esta investigación). Tales documentos se diferencian de la mayoría de los registros bibliográficos habituales por su tamaño y complejidad, al mismo tiempo que difieren de los documentos textuales tradicionales por su naturaleza altamente estructurada.

¹⁰ Lou Burnard. What is SGML and How Does It Help? *Computers and the Humanities*, 1995, vol. 29, p. 42.

¹¹ Eduardo Peis, Felix de Moya. Sgml y servicios de información. *El profesional de la información*, junio 2000, vol. 9, n° 6, p. 4. Estos autores se basan en otro de los teóricos consolidados de SGML, Eric van Herwijnen. *Practical SGML*. 2nd ed. Boston: Kluwer Academic, 1994.

¹² Edward Gaynor. *From MARC to Markup: SGML and Online Library Systems* [documento HTML]. Charlottesville: University of Virginia Library, Special Collections, 8 de mayo de 1996. Disponible en: http://www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html (consultado el 3 de agosto de 1998).

¹³ Eve Maler, Jeanne El Andaloussi. *Developing SGML DTDs: From Text to Model To Markup*. New Jersey: Prentice Hall, 1996, p. 8-9. Esta distinción o tipificación de los lenguajes de marcado es común en todos los autores que trabajan sobre lenguajes de marcado, v. gr., en el ámbito español: Joseba [cont.]

- a) *Marcado de procedimiento o procesado (procedural)*: relativo a la apariencia física o de formato (tipo de letra, tamaño, etc.), restrictivo a cada sistema de procesador de textos o de edición electrónica de documentos, que no tiene ninguna implicación expresa para el contenido del documento.

- b) *Marcado descriptivo o declarativo*: identifica los elementos estructurales de un documento, determinando su estructura lógica. Se describe la estructura y/o el contenido de un documento, más que su apariencia física en pantalla o impreso. Según este balizado descriptivo, el contenido de un documento se separa del estilo de presentación, de igual forma que los elementos dentro de un documento (capítulo, párrafo, nota al pie, bibliografía, etc.) se categorizan a través de códigos que especifican de cuál de estos elementos se trata. Gaynor¹⁴ señala además que el contenido de los documentos marcados descriptivamente puede reutilizarse con distintas finalidades (por ejemplo, la identificación de los elementos que describen un documento desde el punto de vista bibliográfico: título, autor, etc.).

A esta diferenciación podemos darle una explicación puramente lingüística. Al fin y al cabo, los lenguajes de marcas son, ante todo, lenguajes cuya finalidad es la comunicación, en este caso, entre máquinas que codifican y decodifican el mensaje para presentar al usuario la información. Para esta argumentación lingüística trasladaremos los principios del signo lingüístico asociados originalmente, a las palabras, al signo lingüístico asociado al texto, concretamente al documento (textual) electrónico y, más aún —en el terreno de los metadatos en el que se inserta esta digresión— al documento entendido como objeto de información de la Red (DLO). Podemos decir que el marcado de procedimiento es al documento, lo que el

Abaitua. *Material de referencia para un curso de introducción a SGML* [documento HTML]. Bilbao: Universidad de Deusto, rev. 8 de mayo de 1997. Disponible en: <http://sirio.deusto.es/abaitua/konzeptu/sgml/sgml0.htm> (consultado el 3 de marzo de 2001).

¹⁴ Edward Gaynor. *Op. cit.*, http://www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html

significante (plano de la expresión) es a la palabra; y que el marcado descriptivo es al documento, lo que el significado (plano del contenido) a la palabra.

Además del marcado declarativo y de procedimiento, desde los años 80's, se identifican otras formas de etiquetado. Así, Coombs, Renear y DeRose¹⁵ hablan también de marcado de presentación, de puntuación, referencial e incluso de metamarcado, lo que habitualmente conocemos como metalenguajes (SGML y XML¹⁶). En cualquier caso, independientemente del tipo de marcado (de procedimiento, descriptivo, o simplemente el marcado destinado a la puntuación gramatical de un texto, etc.), éste no se inserta al azar en los materiales textuales sino que siempre trata de transmitir un significado. Cuando un autor lo proporciona como parte de la composición del texto, el marcado expresa su intención y tiene un significado vinculado al formato, lo que Sperber-McQueen, Huitfeldt y Renear denominan *performative significance*¹⁷. Sin embargo, cuando el marcado es parte de la transcripción electrónica de un material preexistente refleja la comprensión del texto y, en ese caso, expresa una argumentación sobre el mismo. En el primer supuesto, el marcado es constitutivo del significado; en el segundo es interpretativo. De todas formas —continúan estos autores¹⁸— *el lector (incluyendo el software que procesa documentos marcados) puede utilizar el marcado para hacer inferencias sobre la estructura y propiedades del texto.*

¹⁵ James H. Coombs, Allen H. Renear and Steven J. DeRose. Markup Systems and the Future of Scholarly Text Processing. *Communications of the ACM*, 1987, vol. 30, issue 11, p. 933-947. También disponible en Internet en: <http://www.oasis-open.org/cover/coombs.html>

¹⁶ *Vid. infr.* 5.1.2 y 5.1.3.

¹⁷ C. M. Sperber-McQueen, Claus Huitfeldt, Allen Renear. Meaning and Interpretation of Markup. En: *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (12. 2000. Glasgow). ALLC/ACH 2000 Conference Abstracts*. Glasgow: University of Glasgow, 2000, p. 27.

¹⁸ *Ibid.*

Siguiendo a Maler y El Andaloussi¹⁹ podemos señalar otros criterios de tipificación de los lenguajes de marcas: por un lado, la dependencia del software, por otro, el contexto. Según la dependencia de la aplicación informática, distinguimos entre el marcado específico de un sistema, que trabaja exclusivamente con una aplicación informática particular que crea y formatea los datos del documento y que se limita a plataformas de hardware concretas (p. ej. a sistemas Unix), y el marcado genérico, que es independiente de las características de un sistema. Es importante matizar, empero, que aunque un lenguaje de marcado sea descriptivo, ello no implica necesariamente que sea genérico. Por ejemplo, los estilos en las plantillas de Word son relativamente independientes, sin embargo, sólo se pueden procesar con MS Word. Según el otro criterio de categorización de este tipo de lenguajes, a saber, el contexto, dichas autoras destacan que la mayoría de los lenguajes de marcado son no-contextuales, es decir, que los datos y el marcado tienen lugar en una secuencia de bits (*stream*) en la que *las relaciones particulares de orden y jerarquía no están explícitamente impuestas*²⁰.

Con todo, de forma elemental, podemos decir que un archivo electrónico realizado según un lenguaje de marcas, es un archivo de texto sin compaginación ni elemento tipográfico alguno, que delimita sus elementos con las etiquetas correspondientes que marcan el principio y el final de los mismos, cuya máxima característica es la estructura lógica de la información. En un sentido amplio de esta concepción, podríamos decir que el marcado de documentos es metainformación sobre los mismos, teniendo en cuenta que permite hacer inferencias sobre la estructura y sobre las propiedades del texto. Chessnut²¹ relaciona el marcado de documentos con los metadatos, en el sentido de que el marcado hace accesibles todas

¹⁹ Eve Maler, Jeanne El Andaloussi. *Op. cit.*, p. 10-11.

²⁰ *Ibid.*, p. 11.

²¹ David R. Chesnutt. The Model Editions Partnership: "Smart Text" and Beyond [documento HTML]. *D-Lib Magazine*, July/August 1997, vol. 3, nº 7/8. Disponible en: <http://www.dlib.org/dlib/july97/07chesnutt.html> (consultado el 19 de agosto de 2000).

las adiciones y visiones de metainformación de un documento científico. Nosotros, en cambio, entendemos que el marcado permite definir estructuras de codificación de la información sobre esa propia información, pero no como metadatos en sentido estricto²². Esta idea puede ser avalada con las reflexiones de Maler y El Andaloussi²³ en su manual para la creación de DTDs (*Document Type Definition*) de SGML donde, tras tratar el tema del crecimiento de la información electrónica y el problema de la recuperación de información, plantean que los ficheros electrónicos normalmente omiten la correcta "información sobre la información", datos sobre su estructura organizacional y su significado. SGML —continúan estas autoras— es una tecnología que aporta un marco para proporcionar esta clase de información, de tal forma que se puede maximizar tanto el valor de los documentos electrónicos como la capacidad de gestionarlos y de acceder a ellos. Estas afirmaciones que Maler y El Andaloussi circunscriben al SGML, se pueden generalizar a los lenguajes de marcado, sobre todo, como veremos, al potencial de XML para maximizar el contenido y el procesamiento de la información a través de esquemas de metadatos.

El desarrollo reciente de estas tecnologías Web y de los lenguajes de *markup* ha extendido las capacidades de los metadatos hacia descripciones más ricas y sofisticadas y, por ende, a un acceso mejor a los recursos electrónicos²⁴. Las dos herramientas que proporcionan una estructura y una semántica a la creación de metadatos son, como venimos anticipando, el *Standard Generalized Markup Language* (SGML) y el potente y flexible subconjunto de éste, *eXtensible Markup Language* (XML). Así, muchas de las iniciativas de identificación de metadatos se

²² Cfr. *infr.*

²³ Eve Maler, Jeanne El Andaloussi. *Op. cit.*, p. 3. Obviamente, en el año de publicación de esta monografía (1996), el problema de la edición electrónica masiva que apuntan Maler y El Andaloussi era mucho menor que en la actualidad (cinco años después) pero ya era exponencial el aumento, y previsible la exacerbación del problema.

²⁴ No en vano Hudgins, Agnew y Brown, se refieren a los metalenguajes como la forma de añadir funcionalidad y extensibilidad a los metadatos (*MetaLanguages: increased functionality, extensibility for Metadata*). Jean Hudgins, Grace Agnew, Elizabeth Brown. *Op. cit.*, p. 3.

aproximan cada vez más a SGML y XML, posibilitando diversos modelos, todos ellos incluidos —por sus requisitos funcionales, método de creación y complejidad de designación— en la banda 3 según la tipología de Dempsey y Heery y dan lugar tanto a metadatos basados en el recurso (que corresponderían a la catalogación e identificación) como a metainformación basada en la materia (que implicarían la clasificación o el análisis de contenido), según la distinción de Ahmed y otros especialistas en XML²⁵.

Algunos de los formatos que se fundamentan sobre la base sintáctica de estos metalenguajes son, por ejemplo: EAD (*Encoded Archival Description*), CIMI (registro del *Consortium for the Interchange of Museum Information*), ICPSR SGML *Codebook Initiative* (para la descripción de conjuntos de datos de ciencias sociales), FGDC o las cabeceras TEI (*Text Encoding Initiative*)²⁶.

5.1.1. SGML (*Standard Generalized Markup Language*)

Sin pretender hacer aquí una argumentación histórica, debemos mencionar que los orígenes de SGML se remontan a los trabajos de Charles Goldfarb, Edward Mosher y Raymond Lorie de 1970, quienes desarrollaron el *Generalized Markup Language* (GML) en el seno de IBM, lo que supuso un gran avance en los lenguajes

²⁵ Kal Ahmed, et al. *Op. cit.*, p. 13-14. Como ya hemos precisado, esta tipificación de los metadatos está también consolidada por otros autores como Kashyap y Sheth *Cfr.* 3.5., sin embargo citamos ahora el reciente libro *Professional XML Meta Data* por ser justo XML el metalenguaje de base estructural que permitirá desarrollar y representar también, además de metadatos basados en el recurso (descripción e identificación), metadatos basados en la materia, es decir, aquellos metadatos que representan las materias y sus interrelaciones, como es el caso de las ontologías o los *topic maps*. *Vid.* OIL, XTM.

²⁶ Aunque a lo largo de este capítulo tendremos oportunidad de hablar de algunos de estos formatos, para información más detallada, *Vid.* Lorcan Dempsey and Rachel Heery. *A Review of Metadata...* *Op. cit.* <http://www.ukoln.ac.uk/metadata/desire/overview/overview.rtf> EAD [p. 30-34]; CIMI [p. 20-24]; ICPSR [p. 56-60]; FGDC [p. 40-45]. Sin embargo, el nivel de desactualización de esta revisión de formatos (*Cfr. supr.* Nota 5), hace que no se haga referencia a la migración que han experimentado de SGML a XML, las DTDs (*Document Type Definition*) de estos formatos, en los últimos años.

prescritos de IBM. La principal innovación, según Price²⁷, fue *el uso de un mercado declarativo (descriptivo) en el que las etiquetas eran nombres que indicaban el tipo de elementos del documento, en lugar de verbos que expresaban qué proceso se iba a realizar*. SGML se constituyó como un estándar internacional en 1986 (ISO 8879²⁸) para la descripción de lenguajes de marcas y para el análisis formal de documentos, al que se refiere Price²⁹ como SGML 86. Se basa en un mercado genérico del documento para identificar sus partes lógicas o elementos constitutivos de forma jerárquica. SGML permite que se definan lenguajes de marcas de forma independiente y, por ello, facilita el intercambio y la conservación de recursos electrónicos ampliamente estructurados. De esta afirmación se desprenden importantes nociones que desarrollaremos a continuación y que explican, de alguna manera, que planteemos una reflexión especial en torno al Lenguaje de Marcas Genérico Normalizado (SGML) en este capítulo dedicado a los modelos de metadatos.

Las ideas que queremos resaltar de la norma SGML 86, relacionadas con el propósito de esta investigación son fundamentalmente dos: que permite que se definan lenguajes de marcas (metalenguajes); y que facilita el intercambio de información de recursos electrónicos ampliamente estructurados.

²⁷ Roger Price. Beyond SGML. En: *ACM Conference on Digital Libraries (3. 1998. Pittsburg)*. Ian Witten, Rob Akscyn, Frank M. Shipman, eds. New York: ACM, 1998, p. 172.

²⁸ Para más información sobre la norma ISO 8879:1986, Vid. <http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=16387>

²⁹ Roger Price. *Op. cit.*, p. 173 y ss. La denominación de Price ("SGML 86") nos servirá para referirnos a la norma específica del SGML del año 86, de tal forma que podamos destacar posterior desarrollo. Este autor analiza las posibilidades de las técnicas de marcado de documentos en un mundo informativo *post-SGML 86*, centrándose, sobre todo, en la integración a otros medios aparte del texto. También nosotros, sin negar el valor histórico de la norma, trataremos de revelar en este capítulo ese mismo mundo informativo post SGML 86 protagonizado, desde nuestro punto de vista, por XML.

De manera sencilla, siguiendo a Lou Burnard³⁰, podemos decir que un documento SGML se estructura de la siguiente forma: en un primer nivel, los textos están compuestos por *entidades*, es decir, por conjuntos de símbolos (caracteres o bytes de datos, marcas, gráficos, etc.). En un nivel más alto de abstracción, un texto se compone de la representación de objetos de varias clases, definidos lingüística o funcionalmente. Tales objetos no aparecen aleatoriamente en el texto sino en relación específica con otros objetos; este nivel de descripción implica que los documentos están compuestos de objetos definidos estructuralmente, lo que en SGML se denomina *elementos*. Por último, la gramática que define cómo tales elementos pueden combinarse adecuadamente en una clase particular de textos se conoce como *tipo de documento*.

SGML no es un lenguaje de marcas en sí mismo, sino que es un estándar (ISO 8879) para la descripción de tales lenguajes de marcado. Esto implica que, como ya hemos mencionado, más que un lenguaje es un metalenguaje³¹, en el sentido de que no se trata sólo de un lenguaje individual, sino de un marco o de reglas para la definición de otros lenguajes de marcas. SGML permite la descripción de información estructurada independientemente de cómo se procese gracias a las DTDs (*Document Type Definition*), un *metadocumento*³² que constituye una sintaxis

³⁰ Lou Burnard. What is SGML... *Op. cit.*, p. 50.

³¹ De forma incidental hemos definido en este trabajo el concepto de metalenguaje (siguiendo al DRAE) como: *lenguaje cuando se usa para hablar del lenguaje mismo* (Cfr. 3.1). Nótese el matiz diferente que implica este sustantivo aplicado a SGML o a XML como un lenguaje que describe otros lenguajes de marcas.

Existe un consenso en toda la bibliografía sobre este tema en torno a la calificación de SGML y XML como metalenguajes. Por ejemplo. Edward Gaynor. *Op. cit.*, http://www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html, ó el trabajo de: Roman S. Panchyshyn, France Bouthillier. *Op. cit.*, p. 141; ó Jon Fausey and Keith Shafer. All My Data is in SGML. Now What? *Journal of the American Society for Information Science*, 1997, vol. 7, nº 48, p. 638; entre otros.

³² Berners Lee define una (*un*) DTD como *un metadocumento que contiene información acerca de cómo puede usarse un conjunto dado de etiquetas SGML. En el mundo XML —continúa el inventor de HTML— este papel lo asumirá un esquema* (Vid. *infr.*). Tim Berners Lee. *Tejiendo...* *Op. cit.*, p. 222.

normalizada para definir descripciones de clases de documentos. Las DTDs definen tipos de documentos y sus estructuras, estableciendo los elementos que requiere una clase concreta de ellos, esto es, *cómo debe etiquetarse dicho tipo de documentos*³³. Cada etiquetado SGML es una aplicación independiente, en el sentido de que un documento constituido según una DTD determinada puede reutilizarse de distintas formas sin que esté vinculado a ningún sistema propietario.

Con todo, SGML es un lenguaje maduro para expresar la estructura de los documentos y se ha aceptado como base de proyectos mayores como TEI e importantes lenguajes y metalenguajes (HTML y XML) y estructuras de metadatos embebidas en ellos. SGML es el marco normativo que proporciona las coordenadas de la búsqueda de información sobre el texto completo más precisa y significativa. El marcado textual asegura que la indización no se separe del texto, algo fundamental en el entorno de la información digital donde el texto, el registro de indización y el propio DLO deben mantenerse e interrelacionarse a través del tiempo y a través de la arquitectura distribuida de la Red. Así, los modelos de metadatos formulados y surgidos a raíz del marcado son la forma más coherente de asegurar una indización persistente y eficaz de los documentos electrónicos.

Distintos lenguajes de la familia de SGML se han utilizado y se utilizan todavía para la creación de bibliotecas digitales y en proyectos de edición electrónica de documentos. Los metadatos en el marcado SGML suponen una codificación de determinadas áreas del documento para identificar al propio DLO en la cabecera (como es el caso de la TEI Header, TEIH). También existen DTDs creadas de forma independiente al documento fuente, como por ejemplo la DTD de metadatos del EAD o la propia TEIH que, como veremos, también puede utilizarse como cabecera independiente para la catalogación.

³³ *A DTD describes how a document conforming to it should be marked up...* Jon Fausey and Keith Shafer. *Op. cit.*, p.638.

A pesar de todas estas ventajas estructurales del SGML, de su amplitud y extensibilidad para crear tipos de documentos en un entorno electrónico y para codificar, de una u otra forma, metadatos asociados a ellos, se trata de un lenguaje muy complejo y difícil de manejar, aplicar y visualizar en el entorno Web (que es sin duda, el contexto de las bibliotecas digitales). Normalmente, la solución para utilizar DTDs de SGML en la estructuración y la gestión de bibliotecas digitales es, como apuntamos al hablar de lo que llamábamos SIG-DLOs (v. gr. Hyperion de Sirsi³⁴), la utilización de Sistemas de Gestión de Bases de Datos Relacionales (SGBDR) como Oracle. Un SGBD ofrece beneficios para la independencia del metalenguaje y de los metadatos, en el sentido de que los elementos de un registro se pueden crear, almacenar y manipular en diferentes formatos. A pesar de estas soluciones, son muy pocas las bibliotecas digitales que no han migrado sus DTDs a la flexibilidad de XML, aprovechando la experiencia en la gestión de metadatos en SGML.

5.1.2. De HTML a XML

En los últimos tres o cuatro años (1998-2001) hemos presenciado la producción de ríos de tinta y sobre todo, miles de *bytes* en la WWW, acerca del declive del HTML como lenguaje de marcado fundamental para la Web y de su transición a XML, más flexible que éste y, a la vez, menos complejo que SGML. En este apartado, trataremos de dar cuenta de esta realidad, sobre todo, de las implicaciones que esto tiene para la metainformación, sin ánimo de hacer un estudio exhaustivo sobre el alcance de XML ni una argumentación desmedida sobre el envejecimiento y la inflexibilidad de HTML³⁵.

³⁴ Cfr. 4.3.2.3.

³⁵ Kiely, por ejemplo, se refiere a HTML en estos términos: *una norma envejecida e inflexible*. Don Kiely. XHTML: A Bridge To The Future [documento HTML]. *Information Week Online*, 8 de mayo de 2000. Disponible en: <http://www.informationweek.com/785/language.htm> (consultado el 28 de agosto de 2000).

HTML³⁶ (*HyperText Markup Language*), el lenguaje de marcas que ha permitido de manera protagónica el desarrollo internacional de la WWW, es una aplicación de SGML conforme con el estándar internacional ISO 8879:1986. Sirve para representar tanto elementos de contenido (marcado estructural) como de presentación (marcado de procedimiento) relativos a un documento. El mayor problema de este lenguaje, no obstante, es que ha sido diseñado, más que para representar el contenido, para soportar una presentación consistente, de tal forma que proporciona muy poca ayuda para la recuperación de información en Internet³⁷. Cada día surgen miles de páginas en HTML, pero a pesar de que es, todavía hoy, el lenguaje por excelencia para el acceso público a los documentos electrónicos a través de la WWW, su facilidad de creación, incrementada por las aplicaciones WYSIWYG, y su inconsistente utilización, unida a la generosidad representativa de los navegadores, han hecho de él un lenguaje de formato, en lugar de un código semántico como podría parecer en un principio.

El lenguaje HTML es demasiado superficial en lo que a representación de la información se refiere. Indica a los navegadores cómo deben presentar la información en vez de lo que la información representa o significa. Frente al crecimiento exponencial de la información Web, es preciso contar con un lenguaje que asegure una mayor consistencia, tanto en la producción como en el procesamiento y

³⁶ World Wide Web Consortium. *HTML 4.01 Specification W3C Recommendation...* *Op. cit.*, <http://www.w3.org/TR/html4>

³⁷ No obstante, algunos sistemas de recuperación genéricos han tratado de aprovechar la exigua semántica del HTML para la recuperación. Por ejemplo, Altavista permite buscar en determinadas partes del documento HTML, aprovechando el marcado. Por ejemplo, en la sentencia: `title:"Universidad Carlos III de Madrid"` buscará sólo las páginas que tengan esa secuencia de caracteres en la etiqueta `<TITLE>...</TITLE>` de la cabecera de los documentos que indiza el buscador. Igualmente permite buscar por `link:` y `anchor:`, aprovechando el etiquetado de enlaces `texto` (`link:` buscará en "url" del enlace y `anchor:` en el texto del ancla). No obstante, aunque cada vez existe mayor homologación en las opciones de búsqueda, en la mayoría de los casos a través de la selección de estas en formularios de búsqueda avanzada, éstas no son comunes a todos los motores de búsqueda de la Red y son, de todas formas, insuficientes para una búsqueda precisa. *Cfr.* Capítulo 7.

distribución de la información, que sea más flexible que el conjunto estanco de etiquetas HTML, y que permita discriminar los distintos DLOs para incrementar el control y la recuperación de información. Ante esta realidad, volver al SGML como base normativa y madura para expresar la estructura de los contenidos Web, parecía excesivo y poco manejable. Por ello surge el XML, como *el esperanto*³⁸, con una vocación unificadora para los lenguajes de la Red, o como la *lingua franca*³⁹ de la Era de la Información. Un metalenguaje con una sintaxis específica y un conjunto de reglas bien seleccionadas, entre el formalismo y la complejidad de SGML y la veleidad del HTML, de tal forma que se puedan concebir lenguajes de marcas específicos para estructurar el conocimiento en distintos tipos de recursos electrónicos.

En toda la abundante bibliografía sobre este nuevo lenguaje de marcado, encontramos expresiones metafóricas y premonitorias de la importancia de este metalenguaje en el desarrollo futuro de la WWW. Así por ejemplo son frecuentes las referencias a XML como *la segunda generación de la Web*⁴⁰, *la penúltima revolución de Internet*⁴¹ y sobre todo, la afirmación de que se trata de una tecnología *más allá de HTML*⁴². En la bibliografía específica se utilizan de forma general todas

³⁸ Jon Bosak, Tim Bray. XML and the Second-Generation Web [documento HTML]. *Scientific American*, May 1999. Disponible en: <http://www.sciam.com/1999/0599issue/0599bosak.html> (consultado el 26 de julio de 1999).

³⁹ Dick R. Miller. XML: Libraries' Strategic Opportunity [documento HTML]. *Library Journal NetConnect*, Summer 2000. Disponible en: <http://www.libraryjournal.com/xml.asp> (consultado el 21 de mayo de 2001).

⁴⁰ Jon Bosak, Tim Bray. *Op. cit.*, <http://www.sciam.com/1999/0599issue/0599bosak.html>

⁴¹ XML la penúltima revolución de Internet [documento HTML]. *Infosociedad*, 2 de septiembre de 1998. Disponible en: http://www-5.ibm.com/es/infosociedad/experiencias/tecnologia_9808.htm (consultado el 6 de septiembre de 1998).

⁴² La expresión *XML: Beyond HTML (XML: más allá de HTML)* es recurrente en la bibliografía sobre este tema así por ejemplo la encontramos, entre otros, en: Todd Freter. *XML: Mastering Information on the Web* [documento HTML]. Sun Microsystems, 10 de marzo de 1998. Disponible en: <http://www.sun.com/980310/xml> (consultado el 28 de agosto de 2000); Tim Bray. *Beyond HTML: XML and Automated Web Processing* [documento HTML]. *View Source Magazine*. Netscape [cont.]

estas expresiones para referirse a XML, para captar la atención de usuarios y creadores de páginas Web. Sin embargo, muchos autores como Freter⁴³, sistematizan los problemas de HTML para que surja un lenguaje "más allá" de él, en estos términos:

- HTML es sólo una tecnología de presentación. No refleja necesariamente información sobre los datos a los que se aplican las etiquetas de marcado HTML. Por ejemplo, se sabe que <H2>Apple</H2> tiene una apariencia previsible y concreta en un navegador Web (aparecerá en negrita, como un encabezado de segundo nivel). Sin embargo, no aporta ninguna información semántica sobre el término "apple", no precisa si se trata de una compañía de informática o de una fruta en inglés, sólo indica cómo se visualizará dicho término en el agente de usuario (AU) o navegador.
- HTML tiene un conjunto de etiquetas prefijado y finito⁴⁴. No se pueden ampliar o crear nuevas etiquetas más significativas, sólo el Consorcio Web (W3C) puede

Communications Corporation, 1999. Disponible en: http://developer.netscape.com/viewsource/bray_xml.html, y en: <http://xml.coverpages.org/brayXMLViewsource.html> (consultado el 3 de septiembre de 1999).

⁴³ Todd Freter. *Ibid.*, [http://www.sun.com/980310/xml/;\\$sessionid\\$TUZ5IYAAE5J1AMTA1LU3NQ#HTML'sproblems](http://www.sun.com/980310/xml/;$sessionid$TUZ5IYAAE5J1AMTA1LU3NQ#HTML'sproblems) y [http://www.sun.com/980310/xml/;\\$sessionid\\$TUZ5IYAAE5J1AMTA1LU3NQ#XML'sGoals for the internet](http://www.sun.com/980310/xml/;$sessionid$TUZ5IYAAE5J1AMTA1LU3NQ#XML'sGoals for the internet)

⁴⁴ Por ejemplo, en la versión 3.2 de HTML no se habían incluido las etiquetas relativas a la estructuración Web a través de marcos o *frames*, sin embargo, a partir de la versión 4.0, ya se incluían, dentro de la DTD frameset.dtd, todos los elementos relacionados con la visualización múltiple de documentos en marcos (<FRAMESET>, <FRAME>, <NOFRAME>). Cfr. World Wide Web Consortium. *HTML 3.2 Reference Specification: W3C Recommendation, 14-Jan-1997* [documento HTML]. Dave Raggett, W3C, 14 de enero de 1997. Disponible en: <http://www.w3.org/TR/REC-html32> (consultado el 19 de octubre de 2000); World Wide Web Consortium. *HTML 4.0 Specification: W3C Recommendation, revised on 24-Apr-1998* [documento HTML]. Dave Raggets, Arnaud Le Hors, Ian Jacobs, eds. W3C, 24 de abril de 1998. Disponible en: <http://www.w3.org/TR/1998/REC-html40-19980424> (consultado el 19 de octubre de 2000). Concretamente en la especificación de HTML de 1998 se habla de los *frames* en: <http://www.w3.org/TR/1998/REC-html40-19980424/present/frames.html>

hacerlo, como demuestra la propia evolución histórica de este sencillo lenguaje de marcado.

- Por otra parte, los navegadores Web se ven como plataformas de aplicaciones potenciales, pero la tecnología Java, por ejemplo, necesita claramente más de lo que HTML ofrece. Con HTML como estándar de datos, las aplicaciones basadas en la Web se confían a CGI en el servidor para procesar los datos en las páginas Web, lo que contribuye poderosamente al tráfico en la Red y hace la Web más lenta para muchos usuarios.

A pesar de todo esto, volver al SGML como una norma madura para expresar la estructura de los contenidos Web, aunque es posible como demuestra el trabajo de Price⁴⁵ y muchos sistemas de información o bibliotecas digitales fundados en DTDs específicas de SGML, resulta complejo e inapropiado para la Web, fundamentalmente porque necesita una DTD para ser procesado y éstas son difíciles o, más bien, costosas de definir. Además, contiene demasiados elementos que dificultan su uso⁴⁶ y no requiere la conversión a otro formato para poderse visualizar.

Podríamos resumir que, las debilidades de HTML en la imposibilidad de expresar la semántica y la estructura a través de su etiquetado predefinido y fijo, junto a la complejidad de SGML para expresar dicha semántica y estructura, todo ello unido a la necesidad de contar con un formato común que presente información al hombre y que pueda ser procesado por las máquinas, son las circunstancias que han contribuido a que fructifique el lenguaje XML. Aparejados a XML se encuentran también una gran cantidad de modelos de metainformación vinculados a él que soportan distintos requisitos funcionales (bibliotecas digitales, integración de

⁴⁵ Roger Price. *Op. cit.*, p. 172-181.

⁴⁶ Para una comparación exhaustiva de XML frente a SGML haciendo hincapié en las restricciones del nuevo lenguaje frente al de marcado estándar generalizado, *Vid.* James Clark. *Comparison of SGML and XML. World Wide Web Consortium Note 15-December-1997* [documento HTML]. W3C, 15 de diciembre de 1997. Disponible en: <http://www.w3.org/TR/NOTE-sgml-xml.html> (consultado el 2 de septiembre de 1999).

aplicaciones o recuperación de información Web) y que responden a distintos entornos informativos (redes locales, Intranets corporativas o al universo WWW de Internet).

XML⁴⁷ (*eXtensible Markup Language*) es, como SGML, un metalenguaje que, con una tendencia más universal que el HTML y una modularidad y flexibilidad mayor a la de éste, se erige como un verdadero estándar de intercambio de documentos en la Web que mejora la recuperación, ya que facilita el marcado según el contenido semántico de la información⁴⁸. XML constituye un compromiso entre HTML y SGML para las aplicaciones Internet/Intranet y puede considerarse como una generalización del HTML o como un perfil de aplicación de SGML⁴⁹. El primer borrador de XML⁵⁰ surgió a finales de 1996 y se presentó en la conferencia sobre SGML celebrada en Boston entre el 18 y el 21 de noviembre⁵¹ cuando se

⁴⁷ World Wide Web Consortium. *Extensible Markup Language (XML) 1.0. W3C Recommendation 10-February-1998* [documento HTML]. Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, eds., 10 de febrero de 1998. Disponible en: <http://www.w3.org/TR/1998/REC-xml-19980210> (consultado el 24 de agosto de 1998).

World Wide Web Consortium. *Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation 6 October 2000* [documento XHTML]. Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, eds., 5 de octubre de 2000. Disponible en: <http://www.w3.org/TR/2000/REC-xml-20001006> (consultado el 26 de agosto de 2001).

⁴⁸ Nigel Ward and Andrew Wood. *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA98>

⁴⁹ Leloup y otros muchos autores reconocen esta vinculación del XML con sus predecesores SGML y HTML. Catherine Leloup. *Motores de búsqueda e indexación: entornos cliente servidor, Internet e Intranet*. Barcelona: Ediciones Gestión 2000, 1998, p. 117. No en vano por ejemplo, en la elaboración de la especificación de XML 1.0 estaban implicadas personas vinculadas al trabajo con SGML, entre otros, Charles Goldfarb o Eve Maler, cuyo manual sobre SGML hemos citado en los apartados anteriores (Eve Maler, Jeanne El Andaloussi. *Op. cit.*). Esta autora, que primero trabajó para la empresa ArborText y que ahora trabaja en Sun Microsystems, es en la actualidad coeditora de la segunda edición (2000) de la Recomendación del Consorcio sobre XML.

⁵⁰ World Wide Web Consortium. *Extensible Markup Language (XML): W3C Working Draft 14-Nov-96* [documento HTML]. Tim Bray, C. M. Sperberg-McQueen, eds. W3C, 14 de noviembre de 1996. Disponible en: <http://www.w3.org/TR/WD-xml-961114.html> (consultado el 25 de agosto de 2001).

⁵¹ Para una información detallada sobre los contenidos de la Conferencia de Boston, *Vid.* <http://xml.coverpages.org/SG96progCAT.html>

conmemoraba una década de la norma SGML. Desde sus orígenes, se le ha reconocido como el sucesor de HTML⁵² y ha suscitado el interés⁵³ de informáticos, bibliotecarios y de otros muchos colectivos, como por ejemplo el dedicado al comercio electrónico.

Básicamente los documentos XML han heredado la misma estructura de elementos, atributos, etc. de SGML; sin embargo, pueden usarse con o sin DTD. Siempre deben estar bien formados (ajustándose a las especificidades de las sintaxis XML aunque no respondan a ninguna DTD) y, en el caso de que se usen con una DTD específica, deben ser además válidos. A parte de las definiciones de tipos de documentos, XML puede basarse en esquemas. Un *schema XML es una definición de la estructura de una clase de documentos XML, donde el propio esquema puede estar escrito o no en sintaxis XML*⁵⁴. El lenguaje *XML Schema*⁵⁵ permite definir tanto elementos globales, que deben utilizarse de la misma forma en un documento XML, como elementos locales, que pueden tener un significado particular en un contexto determinado.

Desde que se ha empezado a utilizar XML, la necesidad de una forma normalizada de definir la estructura de los documentos se ha convertido en algo cada vez más importante. Tanto las DTDs tradicionales como los esquemas XML son

⁵² Incluso Bill Gates ha reconocido el cambio fundamental que supone XML en el entorno de la información Web a través de la famosa frase: *XML is a real breakthrough* que se utilizó como eslogan de la conferencia sobre XML de 1998 <<http://www.gca.org/conf/xmlcon98>> (consultado el 21 de agosto de 1998). Posteriormente, numerosos autores, también del ámbito de los metadatos y las bibliotecas digitales han reconocido ese "cambio radical" que supone XML, V. gr. Jean Hudgins, Grace Agnew, Elizabeth Brown. *Op. cit.*, p. 4.

⁵³ Olivier Romieux. François Chahuneau: «Tout le monde a intérêt au développement du XML». *Archimag*, abril 1998, nº 113, p. 16.

⁵⁴ Kal Ahmed, et al. *Op. cit.*, p. 26.

⁵⁵ El esquema (*schema*) de XML es una recomendación del consorcio desde mayo del 2001. Vid. World Wide Web Consortium. *XML Schema Part 0: Primer. W3C Recommendation, 2 May2001* [documento HTML]. David C. Fallside, ed. W3C, 2 de mayo de 2001. Disponible en: <http://www.w3.org/TR/2001/REC-xmlschema-0-20010502> (consultado el 13 de mayo de 2001).

tecnologías que proporcionan la descripción de esas estructuras de los DLOs. Mientras que una DTD en el sentido tradicional tiene ya una gran aceptación en la codificación textual en SGML, los esquemas XML deben considerarse aún como una tecnología en fase de implementación. Los esquemas se consideran como una nueva respuesta al problema de definir lógicamente la estructura, contenido y semántica de los documentos. Existen varios argumentos a favor de los esquemas XML: por ejemplo, la sintaxis, el hecho de estar orientados a objetos, el polimorfismo y la posibilidad de definir nuevos tipos de datos para refinar el sistema de marcado que se utilizará en una biblioteca o servicio de información digital, o para estructurar de una forma más coherente la información. Ioannides pertenece al conjunto de autores que defienden la utilización de esquemas frente a las DTDs, argumentando, entre otras cosas, que *XML aspira a ir más allá del limitado reino de SGML ya que tiene el potencial de tratar no sólo texto sino también otros datos estructurados*⁵⁶. Se necesita sustituir las DTDs por esquemas porque éstos permiten una definición del tipo de datos más detallada y comprensible, sin que ello obste para que los datos sigan siendo extensibles y legibles por los hombres.

En cualquier caso, a pesar de su estado incipiente todavía en la Web, e independientemente de que basemos la estructura de un documento electrónico en un esquema o en una DTD, XML reportará muchas ventajas para el procesamiento automático de la WWW frente a HTML y a SGML, que resumimos a continuación:

- XML implica una arquitectura de la información más abierta y extensible, no se necesita versiones para que puedan funcionar en futuros navegadores. Y se puede

⁵⁶ Demetrios Ioannides. XML an Electronic Publishing: XML Schema Languages: Beyond DTD. *Library HiTech*, 2000, vol. 18, nº 1, p. 10. Estas afirmaciones de Ioannides nos parecen fundamentales ya que podemos confiar a los *schemas* XML el reto del procesamiento de diversos tipos de DLOs que pueden formar parte de una biblioteca digital. Una de las mayores críticas a las DTDs es que no están escritas en sintaxis XML, lo que hace que éstas no puedan procesarse por herramientas por ejemplo, para procesar DOM o SMIL y tampoco tienen integración con los *namespaces* (Vid. Glosario) tan importantes en la creación de metadatos. *Cfr. infr.*

visualizar directamente en el navegador utilizando XSL (*eXtensible Stylesheet Language*), el lenguaje de codificación de estilo aparejado a él.

- Los datos están compuestos de múltiples aplicaciones. La extensibilidad y flexibilidad de este lenguaje permitirá agrupar diversas realidades, desde páginas Web hasta bases de datos. Al estar escrito en ASCII en vez de un formato binario propietario, puede interpretarse por distintos software y permite el intercambio de información entre aplicaciones.
- Al ser un metalenguaje eminentemente jerárquico como SGML, hace posible que los datos que deben expresarse en múltiples niveles, se integren en un mismo fichero⁵⁷.
- Permite asimismo la manipulación y/o gestión de datos y documentos desde el propio cliente Web, de tal forma que dichos clientes puedan ser más autónomos para desarrollar tareas que actualmente se ejecutan en el servidor.
- El concepto de hipertexto, en el seno de XML permitirá una denominación independiente y unívoca de la ubicación de los documentos, enlaces bidireccionales, enlaces que pueden especificarse y gestionarse desde fuera del documento, hiperenlaces múltiples, enlaces agrupados, atributos para los enlaces, etc. a través del Lenguaje de enlaces extensible (XLink⁵⁸).
- En cuanto a los procesos de edición electrónica de documentos, basar un documento maestro en XML implica una gran integración y exportabilidad a otros formatos de publicación (papel, Web, CD-ROM, etc.).

⁵⁷ Lograr esa interrelación jerárquica, por ejemplo, en una base de datos relacional, implicaría relaciones complejas entre tablas.

⁵⁸ Vid. World Wide Web Consortium. *XML Linking Language (XLink) Version 1.0. W3C Recommendation 27 June 2001* [documento HTML]. Steve DeRose, Eve Maler, David Orchard, eds. W3C, 27 de junio de 2001. Disponible en: <http://www.w3.org/TR/2001/REC-xlink-20010627> (consultado el 13 de octubre de 2001).

- En cuanto a la recuperación de la información, los motores de búsqueda adaptados al nuevo lenguaje extensible, devolverán respuestas más adecuadas y precisas, ya que la codificación del contenido Web en XML consigue que la estructura de la información resulte más accesible. Se desarrollarán también búsquedas personalizables y subjetivas para robots y agentes inteligentes.
- Con relación a los metadatos, en el seno del consorcio Web, vinculado al desarrollo de XML y a la propia Web Semántica, se está trabajando en la línea de una mayor consistencia, homogeneidad y amplitud de los identificadores descriptivos y de las descripciones de documentos XML a través de RDF⁵⁹ (*Resource Description Framework*), por contraposición a la parquedad descriptiva de los atributos de la etiqueta <META> del HTML⁶⁰.

A pesar de todo ello, XML no sustituye a HTML —de hecho, podemos pensar todavía en una amplia convivencia de ambos lenguajes para la Web— sino a SGML⁶¹. En este sentido, el W3C ha creado también una "norma de transición" que permite crear un documento XML válido que funciona tanto en agentes de usuario que soportan XML como en los "viejos" navegadores creados para HTML. Esa norma es XHTML (*eXtensible HyperText Markup Language*), que reformula, abarca y extiende HTML, y supone el mayor cambio en los lenguajes de la WWW desde que se publicara, en diciembre del 97, la versión 4.0 del HTML⁶². XHTML proporciona

⁵⁹ Cfr. *infr.* 5.2.3.

⁶⁰ Cfr. *infr.* 5.2.1.1 y Tabla 11 para una comparación de las posibilidades de HTML/XML/SGML para expresar metainformación.

⁶¹ Tim Berners Lee. *Tejiendo... Op. cit.*, p. 149.

⁶² XHTML es una norma del W3C desde enero de 2000. World Wide Web Consortium. *XHTML™ 1.0: The Extensible HyperText Markup Language: A Reformulation of HTML 4 in XML 1.0*. W3C Recommendation 26 January 2000 [documento HTML]. W3C, 26 de enero de 2000. Disponible en: <http://www.w3.org/TR/xhtml1> (consultado el 19 de octubre de 2000). Sin embargo el 4 de octubre de 2001 se presentó un nuevo documento de trabajo (*Working Draft*) que constituirá la segunda edición de esta norma. Vid. <http://www.w3.org/TR/2001/WD-xhtml1-20011004> (consultado el 13 de octubre de 2001).

una sintaxis de XML a documentos con elementos HTML y añade ventajas a este último, en ese proceso de transición de formatos al cual aludíamos antes⁶³, ya que permite extensibilidad, portabilidad y modularidad, al añadir nuevos elementos al HTML sin alterar la DTD en que está basado cada documento y al permitir utilizar distintos espacios de nombre (*namespaces*⁶⁴) en un mismo documento.

Las perspectivas de XML para estructurar, intercambiar y recuperar información electrónica son infinitas, pero en nuestra tesis nos interesa este metalenguaje en cuanto a las implicaciones que tiene para las bibliotecas digitales y, sobre todo, para la organización de su contenido a través de metadatos RDF. Roy Tennant, cita a Maslow para ilustrar las posibilidades de XML en este sentido:

*Abraham Maslow dijo una vez, "Cuando la única herramienta que tienes es un martillo, cualquier problema empieza a parecerse a un clavo". En cuanto se entiende XML y las oportunidades que ofrece para la crear y gestionar servicios y colecciones de bibliotecas digitales, se empiezan a ver clavos por todas partes. No es la única herramienta que tenemos, pero es con mucho, la más útil*⁶⁵.

Coincidimos con Tennant, incluso en la forma tan persuasiva de expresar su convencimiento sobre la utilidad de XML en el marco de las bibliotecas digitales. XML, sin ser la única tecnología disponible para organizar colecciones digitales

⁶³ Calificamos XHTML como "una norma de transición" en el sentido que habla Kiely al decir que HTML es el estándar más extendido (*a pervasive standard*) y XML aún no se ha institucionalizado (*is a becoming standard*). Don Kiely. XHTML: A Bridge To The Future [documento HTML]. *Information Week Online*, 8 de mayo de 2000. Disponible en: <http://www.informationweek.com/785/language.htm> (consultado el 28 de agosto de 2000).

⁶⁴ Vid. Glosario, *namespace*. Cfr. 5.2.3.

⁶⁵ *Abraham Maslow once said, "When the only tool you own is a hammer, every problem begins to resemble a nail." Once you understand XML and the opportunities it offers for creating and managing digital library services and collections, you will begin seeing nails everywhere. It is not the only tool you have, but it is by far the most useful.*
Roy Tennant. XML: The Digital Library Hammer [documento HTML]. *Library Journal, Infotech: Digital libraries*, 15 de marzo de 2001. Disponible en: http://www.libraryjournal.com/articles/infotech/digitallibraries/20010315_17960.asp (consultado el 21 de mayo de 2001).

(recordemos por ejemplo, la adaptación del MARC), sí es la más adecuada, ya que permite una gran flexibilidad y esto es fundamental para un sistema de gestión de metadatos. XML es para los metadatos una cuestión técnica de estructuración, a la vez que una solución de sintaxis. Son muchos los autores que hablan de la necesidad y de la oportunidad de XML como formato de codificación en la gestión bibliotecaria⁶⁶, como *un camino abierto para las bases de datos bibliotecarias*⁶⁷, sobre todo, en las bibliotecas digitales⁶⁸.

Si XML es el estándar *de facto*⁶⁹ del Consorcio Web para la representación del contenido (no de la forma) de la información digital, debe ser, asimismo, la base

⁶⁶ Por ejemplo: Dick R. Miller. *Op. cit.*, <http://www.libraryjournal.com/xml.asp>; Jian Qin. Representation and Organization of Information in the Web Space. *Informing Science*, 2000, vol. 3, nº 2, p. 83-88. También disponible en Internet en: <http://inform.nu/Articles/Vol3/v3n2p83-88.pdf>; Sperberg-McQueen, C. M. *XML and What it will Mean for Libraries* [documento HTML]. Chicago: University of Illinois, 30 de junio de 1998. Disponible en: <http://www.uic.edu/~cmsmcq/talks/teidlf1.html> (consultado el 11 de octubre de 2000).

⁶⁷ Richard Gartner. XML: A Way Ahead for the Library Database? En: *Creating Web-Accessible Databases: Case Studies for Libraries, Museums and Other Nonprofits*. Julie M. Still, ed. Medford: Information Today, 2001, p. 157-171.

⁶⁸ Anderson recopiló en 1999 en un artículo genérico las razones para fundamentar las bibliotecas digitales sobre lenguajes de marcado. Hoy en día, nadie pone en duda que XML debe ser la base estructural de una biblioteca digital, en sentido estricto (donde existe una colección real de documentos digitalizados y almacenados de forma centralizada *Cfr.* Capítulo 8). Martha Anderson. A Tool for Building Digital Libraries [documento HTML]. *D-Lib Magazine*, February 1999, vol. 5, nº 2. Disponible en: <http://www.dlib.org/dlib/february99/02journalreview.html> (consultado el 21 de septiembre de 1999).

Tendremos oportunidad de profundizar, de forma práctica, en la utilización de XML en la constitución de bibliotecas digitales cuando hablemos de la BVC en el capítulo 9; sin embargo, son múltiples las experiencias en este sentido. Ameritan mención algunos modelos de recuperación de información en bibliotecas digitales basados en metadatos XML como por ejemplo la biblioteca digital coreana VINIS (*Virtual Library and Information System*) <<http://mgu.vinis.co.kr>>, que se aproxima a la concepción de biblioteca digital que nosotros defendemos en esta tesis (*Cfr.* capítulos 8 y 9). Hyun-Hee Kim and Chang-Seok Choi. XML: How it Will Be Applied to Digital Library Systems. *The Electronic Library*, 2000, vol. 18, nº 3, p. 183-189.

⁶⁹ La idea de estándares *de facto* es muy amplia en la realidad de la Web. Estas especificaciones del W3C, se convierten en "estándares formales de facto" o, dicho de una forma menos paradójica y pretenciosa: *Especificaciones de Acceso Público* (PAS). *Cfr.* 6.2.

estructural primordial para organizar el contenido de las nuevas bibliotecas digitales formadas por DLOs (contenidos digitales), al mismo tiempo que permite la integración, gracias a su flexibilidad, de otros materiales convencionales.

5.1.3. El intercambio de información de recursos electrónicos ampliamente estructurados.

El desarrollo de las tecnologías vinculadas a la WWW y de los lenguajes de señalización (*markup*) están imponiendo uno de los mayores cambios en el tratamiento de la información primaria y secundaria que obliga a los actores de la cadena informativa (editores y bibliotecarios) a cambiar sus prácticas habituales de procesamiento, así como a una convergencia de ambos mundos a través de la aprehensión de formatos sólidos de intercambio de metainformación.

En los últimos años, la investigación sobre información semi-estructurada se ha centrado en modelos de datos, lenguajes de interrogación, y sistemas de bases de datos. El impulso del XML, como nuevo estándar para la representación de datos y el intercambio de información en la Web, ha provocado un nuevo interés para la investigación en estructuración de la información ya que existen notables semejanzas entre los modelos de datos semi-estructurados a los que hacíamos referencia y el XML⁷⁰. Algunos autores, como Martin⁷¹, califican a esta información como *textual semi-estructurada*, en el sentido de que normalmente la capacidad de interrogación de este tipo de información es inferior a la de los recursos formalmente estructurados, pero el volumen de recursos accesibles en Internet hace impracticable la creación de

⁷⁰ Goldman, McHugh y Widom tratan este tema de la semejanza de XML con los lenguajes semi-estructurados tradicionales a través de un ejemplo de migración de información semi-estructurada a XML aplicado a la base de datos Lore <<http://www-db.stanford.edu/lore>>. Roy Goldman, Jason McHugh and Jennifer Widom. From Semistructured Data to XML. *Markup Languages: Theory and Practice*, Spring 2000, vol. 2, issue 2, p. 153-163.

⁷¹ David Martin. *The Information Broker Project* [documento HTML]. SRI International, 1 de noviembre de 1996. Disponible en: <http://www.ai.sri.com/~martin/papers/broker/techrep/memo/memo.html> (consultado el 4 de febrero de 2001).

información totalmente estructurada. No obstante, el uso de registros estructurados y estándares tiene múltiples ventajas y, por ello, se están realizando esfuerzos para el desarrollo de metadatos semi-estructurados (banda 2⁷²) y estructurados (banda 3). Large, Tedd y Hartley⁷³ argumentan al respecto que el objetivo es desarrollar estructuras de datos que tengan algunas de las ventajas de los datos altamente estructurados pero que puedan crearse sin el conocimiento y aptitud que requiere la creación de bases de datos bibliográficas. Además de todo esto, como señalaba Dovey, la aparición de lenguajes de etiquetado o descripción de documentos basados en XML ha fomentado una nueva escuela estructuralista en la caracterización de los metadatos (*Vid.* 3.2.) y, en general, en el ámbito de la recuperación de la información a la que nos adscribimos con esta tesis.

El desarrollo de formatos de intercambio de información ha sido una constante latente en el mundo bibliotecario. José Antonio Frías señala que estos formatos de intercambio pueden ser *fijos, delimitados, etiquetados o balizados*, y se refiere a todos ellos con esta exposición diacrónica de su uso en el contexto bibliotecario⁷⁴:

Los más antiguos, los formatos fijos, tienen una estructura estática, es decir la dimensión de sus campos está determinada. Estos formatos están cayendo en desuso, ya que no reconocen datos de dimensión variable. Los formatos etiquetados han aportado algo más de flexibilidad al hacer preceder la información por una etiqueta que especifica su naturaleza. Más tarde, los formatos ISO 2709, que datan de los años setenta, permitieron la introducción de una jerarquía simple, de un nivel (subcampos, subelementos, etc.), cuyo ejemplo más significativo en el terreno de la catalogación bibliotecaria son los distintos formatos MARC. Por último, con SGML se pueden definir formatos

⁷² Andrew Large, Lucy A. Tedd and R. J. Hartley. *Op. cit.*, p. 137. *Cfr.* 3.5.

⁷³ *Ibid.*, p.138

⁷⁴ José Antonio Frías. La catalogación de los documentos electrónicos: problemática, propuestas e iniciativas. En: *Sistemas de Información. Balance de 12 años de jornadas. Jornadas españolas de documentación automatizada (5. 1996. Cáceres)*. Cáceres: Universidad de Extremadura, Servicio de Publicaciones; ABADMEX, 1996, vol. 1, p. 435, nota 16.

con jerarquías complejas y dar atributos a los elementos de información que se necesitan para una entrada, una marca, un campo de aplicación, una definición, etc.

En esta reflexión queda patente, en el ámbito de desarrollo de la norma SGML y del uso creciente de XML como reducción de ésta, que la necesidad de facilitar el intercambio de información ha llevado a desarrollar formatos cada vez más estructurados. A ello podemos añadir que las unidades de información han recurrido progresivamente a entornos electrónicos para sus productos y servicios, motivadas por el cambio rápido de los medios de difusión de la información y por la incursión de Internet en el mundo de la gestión documental. Las bibliotecas, como productoras consuetudinarias de metainformación, tienen además que reorientar sus servicios y sus operaciones hacia la información y difusión electrónica, máxime si tenemos en cuenta las bibliotecas digitales constituidas *ex novo* de forma electrónica.

Para responder a estas necesidades, teniendo en cuenta, por un lado, la complejidad y heterogeneidad del espacio de información Web⁷⁵ y, por otro, la dificultad que tienen los actuales sistemas de recuperación de información en Internet (SRII), basados en la indización sobre el texto completo para identificar y representar el contexto de la información, es preciso contar con estructuras adicionales (metainformación) que incrementen el contenido de los recursos y que permitan la creación automatizada *de sustitutos documentales*⁷⁶.

Así, los metalenguajes XML y SGML tienen muchas posibilidades para definir modelos de metadatos para estructurar la información, ya que han dado lugar a toda

⁷⁵ Esta complejidad y heterogeneidad la definiremos como VVV de la WWW, esto es: el Volumen, Variedad y Volatilidad que caracterizan la World Wide Web. *Cfr.* 7.1.

⁷⁶ Nigel Ward and Andrew Wood. *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA98> Vid. Glosario para mejor comprensión de *surrogate/subrogate* que aquí hemos traducido como "sustituto documental".

una familia de lenguajes de marcado o *lenguajes de descripción de documentos*⁷⁷. Algunos autores⁷⁸ han comparado, incluso, la forma de escritura y la estructuración del lenguaje de marcas de SGML con la información que, con sus delimitaciones de campo, se introduce en una base de datos. No obstante, Liora Alschuler⁷⁹ —haciendo eco de esta manida comparación en la literatura relativa al SGML y a sus diversas DTDs, entre el lenguaje de marcas y los sistemas tradicionales de sistemas de manipulación de datos en bases de datos jerárquicas, relacionales y orientadas a objetos— establece una brillante comparación entre el texto estructurado en SGML, por supuesto extrapolable a XML, y los datos estructurados en una base de datos:

SGML modela la estructura del lenguaje natural independientemente de que sea un historial, un informe financiero, un manual técnico o un soneto. No está limitado por una noción de estructura derivada del tratamiento de datos o de las matemáticas.

La importancia de los lenguajes de descripción de documentos (*markup*) para la estructuración y el acceso a la información electrónica se evidencia, aún más, teniendo en cuenta que el formato más extendido para el intercambio de información —el formato MARC— falla, como hemos apuntado ya en este trabajo, por su incapacidad para estructurar la información no bibliográfica o analítica, esencial para evaluar y representar los documentos electrónicos, por su imposibilidad para acceder a diferentes niveles de análisis o niveles de detalle (*granularity*⁸⁰) en una documentación fuertemente jerarquizada como la información Web, y por su falta de adecuación para la manipulación y acceso de diferentes versiones de la información

⁷⁷ Tim Berners-Lee. *Tejiendo...* *Op. cit.*, p. 39.

⁷⁸ Catherine Leloup. *Op. cit.*, p. 43.

⁷⁹ Liora Alschuler. *ABCD...SGML: A User's Guide to Structured Information*. London, etc.: International Thomson Computer Press, 1995, p. 33-34.

⁸⁰ *Vid.* Glosario. *Cfr.* 4.1.

Web⁸¹. La catalogación de todos los recursos incluidos en una página Web, como señalan Panchyshyn y Bouthillier⁸², debe permitir el establecimiento de enlaces jerárquicos que indiquen la fuente del documento y las particularidades de su transformación. Mientras que el formato MARC exigía una fragmentación de componentes relacionados de forma jerárquica para crear referencias únicas, SGML y XML permiten presentar los registros estructurados jerárquicamente.

A tenor de las reflexiones hechas por Gaynor y avaladas por otros autores, queda clara la imposibilidad del formato MARC para proveer la flexibilidad y adaptabilidad que requieren los nuevos y emergentes sistemas bibliotecarios fundamentados en la representación y recuperación de información electrónica. A lo sumo, y para una mejor comprensión de las reflexiones que apuntamos aquí, podemos situar en un mismo plano la ISO 2709 y XML porque ambas son estructuras de codificación e intercambio de elementos de datos, pero no podemos identificar la codificación XML y los formatos MARC en tanto que estos últimos se refieren a elementos. El contenido de un registro XML, verbigracia una cabecera de metainformación en sintaxis XML puede ser asimilable a un conjunto de elementos MARC.

La deficiencia reconocida a los distintos formatos MARC, no implica sin embargo, que todo el trabajo de catalogación automatizada realizado desde los años

⁸¹ La insuficiencia del formato MARC para la representación y recuperación de la información electrónica ha sido señalada por Gaynor a través del estudio de tres proyectos que utilizan distintos modelos de metadatos basados inicialmente en SGML: *Text Encoding Initiative* <<http://www.tei-c.org>>, *Finding Aids for Archival Collections* <<http://sunsite.berkeley.edu/FindingAids>> y *Columbia University Digital Image Access Project* <<http://www.cc.columbia.edu/cu/libraries/indiv/avary/diap.html>>. Edward Gaynor. *Op. cit.*, http://www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html, y los resultados ponen, como hemos señalado, en tela de juicio la utilidad del formato MARC para la representación y recuperación de información electrónica, en definitiva, para la recuperación de documentos entendidos como objetos (DLOs) transmitidos a través de la Web (*Cfr.* 4.1).

⁸² Estos autores hacen eco de las conclusiones obtenidas por Gaynor, (*Ibid.*). Roman S. Panchyshyn, France Bouthillier. *Op. cit.*, p. 141.

setenta y fundamentado en este formato de intercambio bibliográfico vaya a convertirse a SGML o a XML tras un esfuerzo ingente e incalculable en términos de tiempo y coste. Lo que sí podemos deducir, y hemos comentado ya, es que el MARC no será el único formato de codificación de datos bibliográficos en las bibliotecas (tradicionales y/o digitales) del nuevo milenio. Por otra parte, el también deficiente formato HTML, que si bien describe perfectamente, en un nivel de procedimiento, los documentos electrónicos, dista mucho de ser un formato apto para definir su estructura conceptual y avalar la recuperación de información relevante en la Web y en las bibliotecas digitales desarrolladas en ella. Las posibilidades de estructuración de metadatos son mucho más ricas en XML que en HTML.

El panorama del intercambio de información pues, se matiza. El contexto actual de la globalización de la información imposibilita o hace muy difícil la persistencia de un formato propio de una profesión. Desde esta óptica, el formato MARC se presenta como un superviviente en la nueva estructuración de la información electrónica, cuya importancia no es más que el producto de su propia historia, y el formato HTML será simplemente un formato de visualización genérica, y los metadatos de sintaxis HTML estarán al albur de la evolución de los sistemas de búsqueda genérica en la Web. El tratamiento de la información electrónica en bibliotecas digitales plantea y exige las siguientes perspectivas: formatos de datos y estándares genéricos que puedan sustraerse de la diversidad de aplicaciones y áreas profesionales o temáticas; la posibilidad de describir sin ambigüedades los diferentes elementos, lógicos y conceptuales, de un DLO; la integración de elementos de indización; la gestión de información jerárquica ilimitada y la integración de información textual y multimedia.

Por ello se están gestando nuevos modelos de metadatos, basados principalmente en SGML y XML, que posibilitan la representación de la información de manera estructurada y eminentemente flexible, *mapeable*⁸³ e intercambiable entre

⁸³ Vid. Glosario, *mapping*.

diversos formatos y que permiten realizar búsquedas robustas basadas en el contenido estructurado de la información Web. Al universo heterogéneo de información de la WWW se une la diversidad de formatos de representación, donde SGML, como norma internacional, y XML como reducto operativo de ésta, servirán de modelo fundamental para la interoperabilidad entre sistemas⁸⁴.

5.2. Propuestas e iniciativas de metadatos.

A lo largo de los capítulos anteriores hemos hablado de RDF, DC, TEI, GILS, AGLS, etc., sin hacer demasiado hincapié en el alcance y significado de estos modelos de descripción de DLOs. Ahora es el momento de analizar el origen, justificación y funcionamiento de algunos de ellos explicando a continuación, las principales propuestas de metadatos.

En primer lugar, trataremos los esquemas (*schemas*) de propósito general, entendiendo por tales aquellos modelos destinados a la representación y descripción de documentos electrónicos, independientemente de su temática y/o finalidad informativa. Dentro de este primer grupo, analizaremos, de una parte, la estructuración de los documentos, evaluando su riqueza semántica para la representación de información y haciendo hincapié en el HTML como sintaxis, y de otra, el formato *Dublin Core* (DC) como semántica de metadatos de carácter general. En segundo lugar, nos referiremos a los formatos de metainformación específicos de un dominio, especialmente la Iniciativa de Codificación Textual para el mercado (TEI) y descripción (TEI Header) de documentos humanísticos o literarios, así como otros modelos para información no textual, y formatos específicos temáticos. Finalmente trataremos el Marco para la Descripción de Recursos (RDF) que

⁸⁴ Vid. 8.3. donde se analiza esta afirmación en mayor profundidad en el contexto de la globalización de la información y del acceso universal a la misma, tratando el concepto fundamental de interoperabilidad y la perspectiva de la Web Semántica.

constituye una promesa y una realidad para la estandarización del procesamiento e intercambio de metadatos relativos a información Web.

Antes de tratar individualmente cualquier modelo de metadatos, se pueden señalar algunas características comunes a todos ellos. Independientemente del área de conocimiento para la que se utilicen, todos los formatos deben dar cuenta de las particularidades de los recursos de información de la Red para facilitar su identificación y posterior recuperación. Asimismo, todos los esquemas tienen unas constantes comunes: *un número limitado de elementos, un nombre para cada elemento o etiqueta, y un significado asociado a cada elemento*⁸⁵. Coinciden también en tener una semántica descriptiva con información relativa al contenido, ubicación, atributos físicos, tipo de DLO (texto, imagen, mapa, etc.) y forma. Los elementos principales, en la mayoría de los modelos de metadatos, cubren el acceso a los documentos publicados en la WWW, incluyendo elementos como el creador del documento, el título, cuándo y dónde se ha publicado, las materias que trata, etc. Burnett, Bor Ng, y Park⁸⁶ señalan que en todos los esquemas de metadatos hay tres elementos obligatorios: título, autor e identificador. Sin embargo, la casuística de modelos es tan amplia y compleja como su finalidad y aplicación. Por ello, en este capítulo nos centraremos en aquellos más estables y desarrollados o en los más utilizados en la comunidad bibliotecaria "digital".

En la descripción de estos formatos, trataremos de destacar los mismos aspectos que utilizamos en la caracterización de tipos de metadatos en zonas o bandas⁸⁷ (Dempsey y Heery): el campo de aplicación, el método de creación, la complejidad de designación, los requisitos funcionales, etc. Realizaremos también

⁸⁵ Chris Taylor. *Op. cit.*, <http://www.library.uq.edu.au/iad/ctmeta4.html>

⁸⁶ Estos autores presentan un cuadro comparativo de seis esquemas de metadatos (DC, URC, Semantic Header, USMARC, IAFA templates y TEI Header), distinguiendo entre los elementos de metadatos intrínsecos y extrínsecos que veíamos al hablar de la tipología de metadatos (*Vid.* 3.5). Kathleen Burnett, Kwong Bor Ng, Soyeon Park. *Op. cit.*, p. 1213-1214.

⁸⁷ *Cfr.* 3.5 (tabla 8).

una breve revisión de los elementos que los constituyen y de las pautas que existen para su construcción, incidiendo, en cada caso, en los aspectos más significativos del modelo. Por otra parte, evaluaremos su estabilidad a través de la prospección de su desarrollo, de modo que nos permita predecir su institucionalización (de facto o formal) como estándares, tema que discutiremos en el capítulo siguiente.

Es preciso matizar que, hablar de modelos de metainformación es hablar de una semántica para la descripción de recursos, pero también implica hablar de una estructura de codificación o de una forma de almacenamiento. Así, la tabla 11 presenta la comparación de las posibilidades de consignar metadatos en HTML, XML y SGML, que anticipan, de alguna forma, las posibilidades de codificación estructural de los modelos de metadatos que analizaremos a continuación, enlazando con las reflexiones del apartado anterior (5.1.) sobre la estructuración de la información.

HTML	XML	SGML
Sólo lo permite los atributos de la etiqueta <META> que pueden soportar esquemas de metadatos concretos como el DC.	Descripción abierta y personalizable a través de RDF (<i>Resource Description Framework</i>) y de otras tecnologías relacionadas.	Proyectos concretos de codificación de metainformación como parte de DTDs específicas (ej. TEI, EAD).

Tabla 11. Estructuras de datos y metadatos

5.2.1. Metadatos de propósito general

Por modelos de metadatos de propósito general, entendemos aquellos destinados a la representación y descripción de documentos electrónicos que, independientemente de su temática y/o finalidad informativa, proporcionan datos de diverso tipo sobre los recursos. Bajo este epígrafe, hablaremos expresamente de metadatos HTML (metaetiquetas), como modelo sintáctico de definición para almacenar la expresión del contenido de un documento y como metainformación más básica —a través de las propiedades `keywords` y `description`— y del Dublin Core, como modelo semántico concreto para expresar dicho contenido de una forma homogénea, destacando en ambos casos la vocación genérica de los atributos.

5.2.1.1. Estructuras genéricas de metainformación

En un sentido amplio, y partiendo del alcance que le damos aquí a los metadatos de propósito general, podríamos entender el marcado propiamente dicho, esto es, las etiquetas en sí mismas, como metadatos. Ello se debe a que las etiquetas XML pueden entenderse como información adicional que proporciona estructura y nombres a los elementos, es decir, información sobre la información⁸⁸. Por ejemplo:

```
<BOOK>
<TITLE>El Quijote</TITLE>
<AUTHOR>Miguel de Cervantes</AUTHOR>
</BOOK>
```

Sin embargo, un documento XML no es sólo un conjunto de datos de carácter con etiquetas, sino una estructura completa constituida por elementos y contenidos. La combinación de datos de carácter con etiquetas es la forma convencional de representar dicha estructura. Los metadatos y la inclusión directa de metainformación dentro de los documentos XML utilizando la capacidad de sus elementos para autodescribir metadatos, *son enfoques diferentes*, como afirman Martín Galán y Rodríguez Mateos⁸⁹. La utilización de XML puro para la inclusión de

⁸⁸ Muchos autores recogen esta opinión sobre el marcado XML, definiéndolo como "marcado semántico". Rovin Cover, por ejemplo, sostiene que el marcado en sí mismo es una forma de 'metadatos' que nos explica cuáles son los elementos constituyentes (por nombre), y cómo los objetos de información están estructurados en unidades coherentes más amplias (*The markup itself is a form of 'metadata', explaining to us what the constituent elements are (by name), and how these information objects are structured into larger coherent units*). Robin Cover. XML and Semantic Transparency [documento HTML]. En: *The XML Cover Pages*. Oasis, 24 de noviembre de 1998. Disponible en: <http://www.oasis-open.org/cover/xmlAndSemantics.html> (consultado el 16 de octubre de 2000).

Para Kashyap y Sheth XML y RDF son lenguajes complementarios que completan un propósito diferente. *XML proporciona un conjunto de etiquetas "semánticas" para describir los datos que subyacen, y puede utilizarse para expresar descripciones de metadatos. La raíz de RDF es describir los datos que subyacen a un documento pero desde un punto de vista semántico*. Virpul Kashyap, Amit Sheth. *Information Brokering... Op. cit.*, p. 22-23. Por ello, nuestra tesis se basa en la defensa de RDF como una forma de usar XML orientado a datos (a metadatos) en lugar de a documentos, en aras a constituir lenguajes de lógica semántica que puedan ser entendibles por el hombre y por el ordenador. *Cfr. infr.* 5.2.3.

⁸⁹ Bonifacio Martín Galán, David Rodríguez Mateos. Estructuración de la información mediante XML: un nuevo reto para la gestión documental. En: *Jornadas Españolas de Documentación* (7. [cont.]

metainformación (esto es, añadir semántica a la propia estructuración de XML) o la adopción de un modelo de metadatos específico dependerá del propio sistema de recuperación de información y también del tamaño de la colección.

Basados en esa capacidad de autodescribir metadatos del XML, surgieron algunos modelos destinados a la comunidad general de desarrolladores Web o creadores de contenidos para la Web. Así, por ejemplo, en 1997, Apple creó un formato de metadatos denominado **MCF** (*Meta Content Framework*) a partir de los resultados del programa de investigación "proyecto X", cuyo objetivo consistía en crear un formato que proporcionara una estructura extensible para la codificación y el intercambio de metadatos⁹⁰. En junio de ese mismo año, Netscape envió al W3C⁹¹ una propuesta de modelo de datos capaz de describir *estructuras de organización de la información (metadatos) para colecciones de información en red*, utilizando para ello la sintaxis para la representación de esos datos en XML. Esa propuesta era MCF, que entrañaba una concepción de metadatos que luego se desarrollará en otras iniciativas como RDF: crear un marco robusto para la representación de metadatos de todo tipo (información comercial, canales, autoría, descripción, etc.). El modelo de MCF descansaba en etiquetas (tipos de propiedad), nodos y arcos (propiedades) para expresar las relaciones entre las distintas entidades o DLOs. A pesar de que era un formato *a priori* adecuado para las bibliotecas digitales, su implementación, como

2000. Bilbao). *Op. cit.*, p. 119. Como ya dijimos, la capacidad del marcado descriptivo o declarativo (p. ej. XML) para la identificación de los elementos que describen un documento desde el punto de vista bibliográfico, ha sido señalada además por otros autores como Gaynor. Edward Gaynor. *Op. cit.*, http://www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html

⁹⁰ Jon Knight. Will Dublin form the Apple Core?[documento HTML]. *Ariadne*, 23- January-1997, issue 7. Disponible en: <http://www.ariadne.ac.uk/issue7/mcf/intro.html> (consultado el 19 de febrero de 2000). En este artículo, Knight habla de la implementación del Dublin Core dentro de la sintaxis MCF en herramientas propietarias de Apple como *HotSauce*. Sin embargo, como demuestra el desarrollo posterior de los metadatos, el uso del MCF no se generalizó, a favor de otras estructuras como RDF.

⁹¹ *Meta Content Framework Using XML*. Submitted to W3C 6 June 97 [documento HTML]. R.V. Guha, Tim Bray, eds. W3C, 24 de junio de 1997. Disponible en: <http://www.w3.org/TR/NOTE-MCF-XML-970624> (consultado el 25 de noviembre de 1998).

demuestra la falta de bibliografía sobre el tema, ha sido muy escasa, aunque sí han prosperado algunas de las ideas introducidas a través de este modelo, por ejemplo en RDF.

Con objetivos semejantes a MCF, Microsoft planteó su propia iniciativa, también una aplicación XML, denominada *Channel Definition Framework/Format*⁹² (CDF) de igual forma enviado al Consorcio Web como una propuesta en marzo de 1997. En este modelo se entiende un canal (*channel*) como un conjunto de documentos o un grupo de contenido que puede enviar, recibir u operar como una unidad⁹³. Al igual que MCF, es un formato de propósito general, ya que está pensado para la interoperabilidad entre diferentes plataformas y motores de búsqueda para una recuperación genérica en la WWW, y puede entenderse también como un formato estructural de metainformación como RDF. Quizás por ello, y a pesar de su potencial para la estructuración de la información, su desarrollo ha quedado relegado por RDF que sí se ha convertido en una Recomendación del Consorcio Web, como veremos.

El propio **marcado HTML** proporciona metainformación de carácter general, en el sentido de que esos datos sobre los datos serán recuperados por sistemas de búsqueda de información destinados a toda la World Wide Web. HTML, a pesar de haberse convertido en un lenguaje de formato, originalmente es un lenguaje semántico. Así por ejemplo, marcar una determinada secuencia de caracteres con un encabezado de primer nivel (<H1>Metadatos y recuperación de información: estándares, problemas y aplicabilidad en bibliotecas digitales</H1>) implica que esa información es más importante que un encabezado de segundo nivel <H2>, de tercer nivel <H3>, etc. o

⁹² Castedo Ellerman. *Channel Definition Format. Submitted to de W3C on 09 March 1997* [documento HTML]. W3C, 3 de octubre de 1997. Disponible en: <http://www.w3.org/TR/NOTE-CDFsubmit.html> (consultado el 12 de enero de 1999).

⁹³ *Ibid.*

que el texto marcado dentro de un párrafo normal <P>. No obstante, como venimos afirmando, la versatilidad en la utilización de este lenguaje, junto a la permisividad de los agentes de usuario al interpretar el marcado, han hecho que esa semántica implícita en las marcas, inicialmente estructurales, se hayan convertido en una señalización de formato.

Como hemos apuntado anteriormente, HTML tiene su propia sintaxis para la asignación de metainformación a los documentos Web: las etiquetas <META>. Estas metaetiquetas han sido, y siguen siendo —al menos en lo que respecta al dominio general de la WWW— las principales contenedoras de metainformación, a pesar de que, como veremos, sus limitaciones semánticas y la inconsistencia en su uso han dado lugar a distintos esquemas o sistemas de metadatos, bien de propósito general o específico, que tratarán de dar mayor solidez a la recuperación de información Web a través de la codificación y la sintaxis de XML.

Cuando hablamos de "dominio general de la WWW" nos referimos a cualquier página que forma parte de la Web sin que sea un DLO *ad hoc* creado para una biblioteca digital o para un sistema de recuperación de información concreto. A ese dominio general que aludimos pertenece cualquier página publicada en Internet, susceptible de ser recuperada por cualquiera de los SRII de tipo genérico que mencionaremos en el capítulo 7 (esto es, Altavista, Google, etc.). En estas páginas, la utilización de metadatos tiene una finalidad exclusivamente promocional o persigue aumentar la visibilidad en los motores de búsqueda. Están basadas, desde el punto de vista estructural, en metaetiquetas HTML que son las únicas que "entienden" todavía los sistemas genéricos de búsqueda y recuperación de información en Internet⁹⁴. Todo ello no obsta para que las metaetiquetas HTML permitan generar SRII

⁹⁴ Esta misma concepción de "dominio general de la WWW" es la que utilizamos en el capítulo anterior al clasificar las herramientas de creación de metadatos y en la que nos basamos para distinguir las herramientas destinadas a la creación de metaetiquetas, por contraposición a dominios informativos concretos —principalmente, bibliotecas digitales, o contextos finitos de información, según nuestra propia tesis— que justificaban la existencia de herramientas destinadas a la creación de metadatos propiamente dichos. *Cfr.* 4.3.

específicos o bibliotecas digitales basados en una estructuración exclusivamente HTML. El hecho de que le demos tanta importancia a la estructura HTML en este apartado dedicado a los metadatos de propósito general, se debe justamente a que en ella se funda la recuperación de información genérica, uno de los escenarios de recuperación de información en Internet que queremos reflejar en esta tesis, por contraposición al de las bibliotecas digitales.

Aunque en las recomendaciones del W3C sobre HTML (4.0 y 4.01) ya se aludía a la existencia de un marco más amplio para la descripción de recursos (RDF⁹⁵), ambas incluyen la etiqueta <META> como parte estructural de un documento HTML. A diferencia del resto de las etiquetas HTML, el elemento <META> se usa para proporcionar información sobre el sitio o página Web, no sobre el formato de la página, ya que el contenido de estas etiquetas no se visualizará en el agente de usuario por estar en la cabecera del documento.

La información que pueden expresar estas estructuras HTML puede ser de distinta índole y utilidad tanto para los navegadores como para los motores de búsqueda. Estarán formadas por el elemento <META>, una propiedad y un valor para esa propiedad: <META name="propiedad" content="valor"> (Ej.: <META name="author" content="Eva Méndez">). Según esto, el proceso de asociación de metadatos en HTML a una página o recurso Web, implica:

⁹⁵ En la especificación de HTML de 1998, al hablar de la etiqueta <META> incluida en el encabezamiento de las páginas, se dice expresamente que, en el momento de la redacción de dicha especificación, se estaba desarrollando *un lenguaje de descripción de recursos (RDF) que sirva para que los autores puedan asignar información más rica legible por ordenador, para describir sus documentos HTML, así como otros recursos accesibles en la Red*. World Wide Web Consortium. *HTML 4.0 Specification...* Op. cit., <http://www.w3.org/TR/1998/REC-html40-19980424>

Finalmente, RDF sería una recomendación W3C en febrero de 1999. [RDFMS] World Wide Web Consortium. *Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation 22 February 1999* [documento HTML]. Ora Lassila and Ralph R. Swick, eds. W3C, 24 de febrero de 1999. Disponible en: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222> (consultado el 1 de marzo de 1999). Por ello, la recomendación del HTML 4.01 de diciembre de 1999, el W3C remite expresamente al texto sobre RDF. World Wide Web Consortium. *HTML 4.01 Specification...* Op. cit., <http://www.w3.org/TR/REC-html40/struct/global.html#h-7.4.4>

- a) Declarar una propiedad y el valor de esa propiedad, bien dentro del mismo documento, con el elemento <META> en HTML, o bien fuera del documento, a través de un enlace con el elemento <LINK> en el <HEAD>. A veces, el elemento LINK puede utilizarse para especificar metadatos propiamente dichos cuando éstos remiten a un URI. Por ejemplo:

```
<LINK REL="DC.Creator" HREF="http://www.bib.uc3m.es/~mendez">
```

- b) Referir a un perfil donde se definen las propiedades y sus valores legales a través del atributo `profile` del elemento <HEAD>. La declaración del `profile`, tiene una función semejante a la declaración de *namespaces* en XML/RDF. La no-declaración del `profile` en la cabecera del documento puede intercambiarse por la legitimación de cada una de las propiedades a través del elemento LINK⁹⁶.

Por ejemplo:

```
<meta name="DC.title" content="metadatos y recuperación de
información: estándares, problemas y aplicabilidad en
bibliotecas digitales"
<lin rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
```

Analizaremos a continuación la idoneidad de esta estructura HTML para representar metainformación a través de los atributos que puede llevar su elemento <META>. Con ello se justifica por un lado, que sea la estructura de metainformación básica que utilizan las páginas y servidores para aumentar su visibilidad en la WWW a través de los motores de búsqueda, y por otro, que su poca flexibilidad para

⁹⁶ La utilización del elemento LINK en HTML, así como la asignación de un perfil, sobre todo cuando la semántica que se codifica en HTML es el modelo DC es, en nuestra opinión, una visión "primitiva" de lo que serán los *namespaces* en XML y RDF, porque en realidad lo que hace el elemento LINK es identificar el espacio del nombre. Vamos a presentar una gradación HTML→XHTML→RDF de la forma que progresivamente se ha ido adoptando para declarar el espacio de nombre (Cfr. *infr.* 5.2.1.2. y 5.2.3. Vid. Glosario, *namespace*):

HTML	<meta name=DC.Subject content="information retrieval"> <link rel="schema.DC" href="http://purl.org/dc/elements/1.1/">
XHTML	<html DC:xmlns=" http://purl.org/dc/elements/1.1/"> <DC:meta name"subject" content="information retrieval"/>
RDF	<RDF DC:xmlns="http://purl.org/dc/elements/1.1/"> <DC:Subject>information retrieval </DC:Subject>

autorizar bibliotecas digitales, requiere modelos más ricos semánticamente como XML. Estos atributos son:

- NAME. Sirve para identificar el tipo de propiedad. El hecho de que la recomendación de HTML no contemple los valores que puede adoptar el atributo name, es la base para que HTML, además de ser el más extendido, adoptado y adaptado para la WWW, sea idóneo para expresar metadatos de propósito general (como por ejemplo, los valores `keywords` y `description` que soportan algunos sistemas de recuperación genérica). Esta adecuación para representar información de propósito general, no obsta para que, bajo una estructura HTML se introduzcan metadatos específicos, que sirvan en teoría, de base estructural de una biblioteca digital.
- CONTENT. Sirve para expresar el valor de la propiedad, bien del atributo name o del atributo `http-equiv`. En este caso, la falta de normalización de los valores que puede adoptar este atributo conlleva el riesgo de *spamming*⁹⁷ cuando se dirige el uso de metadatos a motores de búsqueda de carácter general (SRII), e invalida el carácter documental de estas estructuras para constituir bibliotecas digitales que, en su caso, requerirán de estándares de contenido y normas adicionales para homogeneizar las descripciones.
- SCHEME⁹⁸. Sirve para denominar el esquema que se debe utilizar para interpretar el valor de una propiedad. La utilización del atributo `scheme` dota de cierta "calidad" a las descripciones de metadatos basadas en HTML. Desde la

⁹⁷ Vid. Glosario. Cfr. 7.3.

⁹⁸ El significado de *scheme*, como ya hemos anticipado en el capítulo 4, es complejo en tanto se puede confundir con *schema*. En el contexto HTML, el concepto de *schema* (declarado a través de un *namespace*, Cfr. Glosario) puede asimilarse al elemento `profil` ya que define un esquema de estructura, mientras que el de *scheme* se puede asimilar a un esquema de contenido, vocabulario controlado o lista de autoridad.

recomendación del HTML 4.0⁹⁹ se pueden recoger, a través del atributo `scheme`, por ejemplo, los "calificadores de Canberra" instados por el *Workshop* del Dublin Core para promover descripciones más completas en vez de más extensas, embebidas en la etiqueta `<META>` del HTML¹⁰⁰. La alternativa de representación de esas descripciones más cualificadas sería la utilización, en cada una de las etiquetas `<META>`, del elemento `LINK` aparejado. Las dos representaciones, aplicadas por ejemplo a la descripción de la fecha de publicación de un documento, serían:

- a) `<META NAME="DC.date" CONTENT="2001-07-16">`
`<LINK REL="ISO8601:1988" HREF="http://www.w3.org/TR/NOTE-datetime">`
- b) `<META NAME="DC.date" CONTENT="2001-07-16"`
`SCHEME="ISO8601:1988">`

- `HTTP-EQUIV`. Es un atributo que se utiliza en vez de `name` y permite a los servidores que funcionan con el protocolo de transferencia de hipertexto (HTTP) recabar información de interés para los encabezados del mensaje de respuesta. *V. gr.:*

```
<META http-equiv="Expires" content="30 Nov 2001 14:00:00 GMT">
```

Es útil, por ejemplo, para el trabajo en memoria caché, ya que le indica cuándo debe actualizarse un documento almacenado en ella. Además, este atributo puede tener otros usos, como por ejemplo, la valoración del contenido de la página a través de las especificaciones del modelo PICS (*Platform for Internet Content Selection*), cuando este atributo adopta el valor `"PICS-Label"`.

- Por último, también puede llevar los mismos atributos genéricos que casi todos los elementos HTML que sirven para declarar el idioma de la información (`lang`) y la dirección del texto (`dir`).

⁹⁹ World Wide Web Consortium. *HTML 4.0 Specification...* *Op. cit.*, <http://www.w3.org/TR/1998/REC-html40-19980424>

¹⁰⁰ *Cfr.* Capítulo 4, nota 190. *Vid. infr.* 5.2.1.2

Con todo, y a pesar de que supone la estructura y sintaxis de metadatos más conocida y utilizada de forma general¹⁰¹, la ambigüedad y amplitud de las metaetiquetas en HTML justifica, en cierta medida, la laxitud de su uso y avala tácitamente algunas de las hipótesis sobre las que reflexionamos en esta tesis, tales como la falta de operatividad de los metadatos o metaetiquetas de este tipo en las páginas Web, su escasa utilidad para fundamentar un sistema de recuperación genérico o motor de búsqueda debido justamente a que no son etiquetas obligatorias, así como la necesidad de contar con estructuras flexibles como XML/RDF y esquemas (*schemas*) más robustos que sirvan para crear verdaderas bibliotecas digitales y la interoperabilidad entre éstas.

El W3C reconoce todas estas circunstancias que acontecen al lenguaje de marcado que ha llevado a la WWW a lo que hoy conocemos, y en la especificación de HTML 4.01, de igual forma que relega los aspectos relativos a la presentación de documentos a las CSS, remite a RDF la creación de metadatos. Igualmente se refiere a la necesidad de contar con *schemas* o perfiles de metainformación, mencionando expresamente al Dublin Core¹⁰², formato del que hablaremos a continuación.

¹⁰¹ Así, la corriente más popular de los metadatos y la mayor parte de los estudios que hablan del incremento de las prestaciones de las búsquedas en la WWW a través de metainformación, o que analizan la utilización genérica de metadatos en sitios Web, normalmente se refieren a metadatos codificados en HTML. Algunos ejemplos que corroboran esta afirmación son: José María Angós Ullate, José Antonio Salvador Oliván, María Jesús Fernández Ruíz. Metadatos en documentos HTML: una ayuda para la recuperación de información. *Scire*, Julio-diciembre 1999, vol. 5, nº 2, p. 47-61; Francisco Javier Vidal Bordés, José Antonio Salvador Oliván. *Op. cit.*; José Antonio Merlo Vega, Ángela Sorli Rojo. *Op. cit.*, etc.

¹⁰² El DC es un *esquema (schema)* de metainformación cuyos elementos sirven para nominar las propiedades de un documento electrónico. La sintaxis de codificación puede ser, como veremos, HTML, XML o RDF. La forma de almacenar metadatos DC en la sintaxis HTML está definida en el Request For Comments 2731. Vid. J. Kunze. *Encoding Dublin Core Metadata in HTML: RFC 2731* [documento TXT]. IETF, Network Working Group, December 1999. Disponible en: <http://www.ietf.org/rfc/rfc2731.txt> (consultado el 11 de agosto de 2000).

5.2.1.2. DCMI (*Dublin Core Metadata Initiative*)

Resulta difícil resumir todas las acciones, implicaciones y evolución de la iniciativa de metadatos del Dublin Core, ya que es el formato más divulgado y más aceptado. No sería desproporcionado decir que, independientemente de las críticas que se le puedan hacer, es el estándar más extendido para la representación y recuperación de la información Web, el que más influencia ha ejercido en lo que respecta al desarrollo general del concepto y teoría de los metadatos¹⁰³ y el que mejor se ha adaptado a la evolución de la World Wide Web.

Concebido originalmente para la descripción de recursos Web a cargo del propio autor del DLO, se ha convertido en foco de atención de comunidades involucradas en la descripción formal de recursos (museos, bibliotecas, agencias gubernamentales e incluso organizaciones comerciales).

La Dublin Core Metadata Initiative (DCMI) o simplemente Dublin Core (DC) es un esfuerzo internacional¹⁰⁴ e interdisciplinar dirigido a definir un conjunto básico de elementos para la recuperación de información en Internet (*resource discovery*¹⁰⁵). También es el conjunto propiamente dicho de los (actualmente) quince elementos de datos que puede usarse para describir una amplia variedad de objetos de información

¹⁰³ Por ejemplo, al trabajo en torno al DC se deben conceptos tan importantes en este campo, como el de DLO (*Document Like Object*). Cfr. 3.3., Vid. Glosario, DLO.

¹⁰⁴ Al hablar de que es una iniciativa internacional, nos vemos obligados a hacer una referencia a España donde, de forma más o menos institucional, sólo se ha trabajado con el DC en el grupo iris-index de RedIris http://www.rediris.es/metadata/dublin_core_elements.es.html. Recientemente (17 de mayo de 2001), se ha formado un grupo de trabajo en el marco de SEDIC <http://www.sedic.es/grupo_internet.htm> para la divulgación de las normas de la DCMI y cualquier otra iniciativa relacionada con la normalización en Internet que sea de interés para las bibliotecas y centros de documentación. También hay que destacar la traducción del sitio Web de la DCMI que está realizado dicho grupo de SEDIC <<http://www.dublincore.sedic.es>>, así como la traducción al catalán del DCMES 1.1 <<http://www.gencat.es/bc/dces.htm>>.

¹⁰⁵ Vid. Glosario, *discovery*.

en Internet, con el propósito general de recuperación de recursos multidisciplinares de la Red a través de las siguientes actividades¹⁰⁶:

1. Desarrollar estándares de metadatos para la recuperación de información en Internet a través de distintos dominios.
2. Definir el marco para la interoperabilidad¹⁰⁷ entre conjuntos de metadatos
3. Facilitar el desarrollo de conjuntos de metadatos específicos de una disciplina o comunidad que trabaja dentro del marco de la recuperación de información.

El Dublin Core se diseñó inicialmente para ayudar a los motores de búsqueda a encontrar y recuperar páginas Web —de ahí que lo clasifiquemos como un modelo de propósito general¹⁰⁸— y desde entonces ha ido evolucionando hacia un formato de registro para el intercambio y recuperación de información en el espacio digital.

Este formato recibe el nombre de Dublin Core porque fue en Dublin, en marzo de 1995, donde se celebró el primer encuentro sobre metadatos¹⁰⁹ organizado de forma conjunta por el NCSA (*National Center for Supercomputing Applications*) y OCLC (*On Line Library Computer Center*), donde participaron representantes de la

¹⁰⁶ Stuart Weibel, Traugott Koch. The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions [documento HTML]. *D-Lib Magazine*, December 2000, vol. 6, n° 12, Disponible en: <http://www.dlib.org/dlib/december00/weibel/12weibel.html> (consultado el 23 de febrero de 2001).

¹⁰⁷ Sobre el concepto de interoperabilidad hablaremos en el capítulo 8, al expresar la necesidad de intercambiar información en contextos de información global en Red. *Cfr.* 8.3.2, *Vid.* Glosario, *interoperability, crosswalk/crossroad*.

¹⁰⁸ Por contraposición a otros formatos destinados a la descripción de un tipo de información digital concreta de los que hablaremos posteriormente. *Cfr. infr.* 5.2.2.

¹⁰⁹ El informe de este encuentro fundacional del *Dublin Core* se puede consultar en: Stuart Weibel, et al. *OCLC/NCSA Metadata Workshop Report* [documento HTML]. OCLC, marzo de 1995. Disponible en: http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html (consultado el 15 de junio de 1998).

IETF (*Internet Engineering Task Force*) y expertos en biblioteconomía, informática y codificación textual. El único objetivo consistió en establecer un conjunto central o principal (*core set*) de elementos de metadatos —en un principio iban a ser trece elementos— para facilitar la búsqueda y recuperación de información en la Red. Desde entonces y hasta la actualidad (2001), el interés por este modelo ha ido creciendo y se ha desarrollado a través de sucesivos encuentros o talleres que han servido para impulsarlo. El interés multidisciplinar de este esquema de metadatos y el hecho de que su gestación ha ido a la par de las nuevas estructuras de información de la Web (XML/RDF), han hecho que el conjunto de elementos del vocabulario de Dublin (DCMES, *Dublin Core Metadata Element Set*), se haya convertido recientemente en un estándar formal ANSI/NISO Z39.85-2001¹¹⁰.

Después de todas las modificaciones que se han discutido y acordado a lo largo de la trayectoria del modelo DC, los elementos que constituyen este formato descienden a tres grupos que indican o evocan aproximadamente, la clase o alcance de la información incluida en ellos, y que responden, en cierta medida, a las expectativas que tiene el usuario cuando se enfrenta a la información de la Red¹¹¹. Esta particularidad es muy importante desde el punto de vista de la propia finalidad de los metadatos, ya que dichos elementos pueden establecerse como campos de

¹¹⁰ ANSI/NISO Z39.85-2001: *The Dublin Core Metadata Element Set: An American National Standard Developed by the National Information Standards Organization Approved September 10, 2001 by the American National Standards Institute*. [documento PDF]. Bethesda, Maryland: NISO, rev. 12 de octubre de 2001. Disponible en: <http://www.niso.org/standards/resources/Z39-85.pdf> (consultado el 13 de octubre de 2001). Sobre la formalización de estándares de metadatos, aludiendo específicamente al caso del DCMES, hablaremos en el capítulo siguiente. *Vid.* 6.3.

¹¹¹Al hablar de las aplicaciones de los metadatos en el tercer capítulo de este trabajo (*Cfr.* 3.4), mencionamos las necesidades que tenían los distintos tipos de usuarios al enfrentarse al mundo heterogéneo de la Web, identificadas por Dempsey y Heery —Lorcan Dempsey and Rachel Heery. *Metadata: a current view... Op. cit.* p. 148—. El formato DC cumple, en cierta medida, tales expectativas en lo que se refiere a un usuario como investigador y proveedor de información, sin embargo adolece de elementos que informen a un usuario como cliente y consumidor. Nótese además, la falta de campos que cualifiquen los recursos en cuanto a una valoración de su contenido, por ejemplo para su acceso por menores.

búsqueda en sistemas de recuperación basados en el Dublin Core. Los tres grupos de elementos a los que nos referimos son los siguientes:

- Elementos principalmente relacionados con el contenido del DLO. A este grupo pertenecen: título, materia, descripción, fuente, lengua, relación y cobertura/alcance.
- Elementos relacionados con la propiedad intelectual del recurso. Son las etiquetas relativas al creador/autor, editor, otros autores/colaboradores y derechos de autor.
- Elementos relacionados con la temporalidad del documento. Estos elementos son: fecha, tipo de recurso, formato e identificador del recurso.

El significado concreto de cada uno de los elementos que conforman el modelo del DCMI, se resumen en la tabla siguiente, donde se distinguen estos tres grupos de etiquetas y se precisan las especificidades y la actividad del DCMI en relación con cada uno de los elementos de la versión 1.1. del DCMI (Tabla 12).

ELEMENTO - DC	DESCRIPCIÓN y USO
DC.Title	Nombre dado a un documento, usualmente por el autor.
DC.Subject	Materias del recurso. Se expresan los descriptores o frases que describen el contenido del DLO. Se fomentará el uso de vocabularios controlados y de sistemas de clasificación (<i>schemes</i>).
DC.Description	Una descripción textual del recurso, tal como un resumen, en el caso de un documento, o una descripción del contenido, en el caso de un documento visual.
DC.Source	Secuencia de caracteres utilizada para identificar unívocamente un trabajo a partir del cual proviene el recurso actual ¹¹² .

¹¹² Por ejemplo, es posible usar DC.Source con la fecha de 1607 como descripción de una película basada en una obra de Cervantes, pero es preferible usar el elemento DC.Relation con el calificador "IsBasedOn" con una referencia a un recurso distinto cuya descripción contenga el elemento DC.date con el valor 1607.

ELEMENTO - DC	DESCRIPCIÓN y USO
DC.Language	Lengua/s del contenido intelectual del DLO. El contenido de este campo debería coincidir con la norma ISO 639 ¹¹³ .
DC.Relation	Identificador de un recurso y de la relación que tiene con el objeto de información que se está describiendo ¹¹⁴ .
DC.Coverage	Característica de cobertura espacial y/o temporal del contenido intelectual del recurso ¹¹⁵ .
DC.Creator	Persona física o jurídica responsable de la creación del contenido intelectual del recurso. P. ej., los autores, en el caso de documentos escritos, artistas, fotógrafos e ilustradores en el caso de recursos visuales.
DC.Publisher	Entidad responsable de hacer que el recurso se encuentre disponible en la Red en su formato actual. (p. ej., la empresa editora, una universidad u otro tipo de organización).
DC.Contributor	Persona u organización que haya tenido una contribución intelectual significativa en la creación del documento (p. ej., editor, ilustrador, traductor, etc. en la concepción ISBD tradicional).
DC.Rights	Referencia (URL, por ejemplo) sobre derechos de autor, bien a un servicio de gestión de derechos o a un servicio que dará información sobre los términos y condiciones de acceso a un recurso electrónico o DLO.

¹¹³ RFC 1766 Tags for the Identification of Languages (ISO 639) <<http://www.ietf.org/rfc/rfc1766.txt>>.

¹¹⁴ Este elemento permite enlazar los recursos relacionados y las descripciones de los recursos. Por ejemplo: IsVersionOf (versión de un trabajo), IsBasedOn (p. ej., la traducción de un documento), IsPartOf (p. ej. el capítulo de un libro), IsFormatOf (transformación de un conjunto de datos en una imagen). Para asegurar la interoperabilidad, estas relaciones deben seleccionarse de la lista aprobada después de la conferencia de Frankfurt: Dublin Core Qualifiers (DCMI Recommendation, 11 de julio de 2000) <<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers>>.

¹¹⁵ La cobertura espacial se refiere a una región física, uso de coordenadas (longitud y latitud) o nombres de lugares, preferiblemente extraídos de una lista controlada o *scheme*. La cobertura temporal se refiere a la temporalidad del contenido del recurso, no a la fecha de su creación o publicación (que se consignará en el elemento DC.Date.), pero usa el mismo formato ISO 8601/ W3CDTF. Cfr. *infr*.

Existe un grupo dentro del DCMI que, a través de un documento de trabajo (*Working Draft*) del 8 de abril de 1999, trata de revisar la RFC 2413: *Dublin Core Metadata for Resource Discovery* <<http://www.ietf.org/rfc/rfc2413.txt>> y de precisar el alcance y utilización del elemento Coverage conforme a la ISO 11179 sobre registros de metadatos.

ELEMENTO - DC	DESCRIPCIÓN y USO
DC.Date	Fecha en la que el recurso se puso a disposición de los usuarios en su forma (electrónica) actual ¹¹⁶ .
DC.Type	Categoría del DLO en cuanto al tipo de información que representa (p. ej., imagen, conjunto de datos, texto, software, etc.).
DC.Format	Formato de datos de un DLO, usado para identificar el software y, posiblemente, también el hardware que se necesita para mostrarlo ¹¹⁷ .
DC.Identifier	Secuencia de caracteres usados para identificar unívocamente un recurso. Para los recursos en línea, estos identificadores serán URIs (URL/URN) ¹¹⁸ . Para otros recursos pueden usarse otros formatos de identificadores, como ISBN, ISSN, etc.

Tabla 12. Elementos del modelo Dublin Core

La mejor forma de analizar e interpretar este modelo de metadatos es revisar su desarrollo, importancia y evolución a través de los distintos encuentros, talleres o grupos de trabajo. Por ello presentamos un breve resumen de lo más característico de cada uno de ellos, reconociendo que cada taller o conferencia del DCMI podría

¹¹⁶ En correlación con lo apuntado en la nota anterior, esta fecha no debe confundirse con la del elemento *DC.Coverage* que se asocia a un recurso sólo en la medida en que el contenido intelectual del mismo está relacionado con esa fecha. Para determinar la fecha de publicación o de disposición al público se utiliza el formato del Consorcio Web, *Data and Time Format (W3CDTF)* <<http://www.w3.org/TR/NOTE-datetime>> basado en la norma ISO 8601 que incluye, entre otras, el formato de fecha AAAA y AAAA-MM-DD.

¹¹⁷ Para asegurar la interoperabilidad, los valores, tanto del elemento *DC.Format*, como del *DC.Type* deberían seleccionarse de entre una lista de valores. Sobre al elemento *type*, existe una recomendación, *DCMI Type Vocabulary* <<http://www.dublincore.org/documents/2000/07/11/dcmi-type-vocabulary>>, desde julio de 2000. En cuanto al elemento *format*, se ha creado un grupo de trabajo (*Format Working Group* <<http://www.ukoln.ac.uk/metadata/resources/dc/dc-format-wg/review2413.html>> con el mismo objetivo que el grupo destinado al elemento *Coverage*: adaptar, en este caso el elemento *format*, a la representación ISO 11179.

¹¹⁸ Respecto a la identificación de recursos, es preciso recordar las iniciativas que se están gestando sobre este tema en el seno del IETF (URN, URC) y otras tentativas individuales (DOI, WEBDAV, OpenURL, etc.), todas ellas válidas para identificar de manera uniforme un DLO. *Cfr.* Capítulo 4, nota 12.

generar una larga digresión o discusión que rebasaría el objetivo de este capítulo¹¹⁹. De esta forma, nos limitaremos a exponer brevemente lo más significativo del *iter* del DC a través de los *workshops*¹²⁰ celebrados en torno a él. En cada uno de los talleres o conferencias se marcan los avances, así como los objetivos futuros de este modelo de metainformación.

- **DC-1** (marzo 1995. Dublín, Ohio. OCLC/NCSA): Partiendo de la dimensión creciente de la información existente en Internet y de la imposibilidad de que los catalogadores, indizadores y analizadores de información la trataran con los sistemas existentes hasta entonces, en este primer encuentro se llegó al acuerdo de constituir una alternativa basada en un modelo de metadatos constituido por 13 elementos (materia, título, autor, editor, colaboradores, fecha, tipo de objeto, formato, identificador, relación, fuente, lenguaje y cobertura) de tal forma que los propios autores o editores de la información Web pudieran describir sus recursos. Este primer encuentro se centró en la descripción semántica dirigida expresamente al problema de la recuperación de información genérica de documentos electrónicos.
- **DC-2** (marzo 1996. Warwick, Reino Unido). UKOLN/OCLC): En el segundo encuentro del Dublin Core, se amplió el consenso a la comunidad internacional y se mencionó el problema de la sintaxis para las aplicaciones basadas en la Web. El avance más importante de esta reunión fue el establecimiento de un modelo conceptual para cualquier arquitectura de metadatos que se denominó, por haber

¹¹⁹ Para profundizar en el contenido de cada uno de los talleres del DC, se pueden consultar los distintos informes realizados por Stuart Weibel a tenor de ellos y publicados en distintas revistas profesionales (*D-Lib Magazine*, *Bulletin of ASIS*, etc.), que suelen ser una divulgación formalizada del *statu quo* del DC en cada momento. (Cfr. Bibliografía). Asimismo, se puede acceder a las ponencias y contenidos de cada uno de ellos en la propia página de la DCMI <<http://dublincore.org/workshops>>.

¹²⁰ Aunque hemos mencionado en diversos puntos de este trabajo el término *workshop*, evocando el concepto de taller o reunión de trabajo formal que aúna a especialistas en un tema (en este caso, en un modelo de metadatos), para una visión más profunda de este término, así como de su utilización en el contexto de esta investigación, *Vid.* Glosario.

surgido de ella, *Warwick Framework* (marco de Warwick), el cual sería entendido como la *arquitectura conceptual que actúa como 'contenedor' de distintos paquetes de metadatos heterogéneos*. Este principio de organización que surgió del segundo *workshop*, se basa en el reconocimiento de la existencia de una gran variedad de modelos de metadatos, desarrollados de forma independiente y gestionados por diferentes comunidades según distintos formatos. Ante esa realidad, el taller de Warwick advirtió la necesidad de que los diferentes formatos convivieran en la Red, lo que requería una única arquitectura que permitiera la evolución coherente de los diversos componentes de los metadatos. Este marco, junto con el *Meta Content Framework* (MCF), forman el núcleo del desarrollo del RDF que facilitará un modelo sintáctico flexible para los metadatos basados en la Web (*Vid. infr.*).

- **DC-3** (septiembre 1996. Dublín, Ohio. *OCLC/CNI —Coalition for Networked Information—*): El tercer taller del DC, al que asistieron principalmente expertos en imagen, amplió la cobertura de recursos en los que se centraba el DC (hasta entonces páginas Web e información textual) a las imágenes. Se partió de la base de que, si bien el texto difiere mucho de la imagen, las categorías para describirlas no son tan diferentes, de modo que el formato del DC sirve para la descripción de un gran número de recursos de imagen, especialmente aquellos que comparten características con otros DLOs. Para reafirmar esta adecuación del modelo-DC a la descripción de imágenes, se añaden dos elementos más al formato: se separa un campo de descripción (*DC.Description*) del de materia (*DC.Subject*), se agrega el campo de derechos (*DC.Rights*) y se revisa también la definición de algunos de los elementos existentes.
- **DC-4** (marzo 1997. Canberra, Australia): Del taller de Canberra emerge, sobre todo, un amplio espectro de filosofías que podemos resumir en dos tendencias: una minimalista y otra estructuralista, que vienen a ser algo así como el conceptismo y el culteranismo del Dublin Core en cuanto a su complejidad

estructural¹²¹. En la tendencia minimalista se valora la simplicidad partiendo de que la aceptación y la interoperabilidad pueden decaer vertiginosamente si se incrementa la complejidad de los estándares; en la corriente estructuralista, en cambio, si bien se reconoce ese riesgo, se considera que la capacidad para calificar el contenido de los metadatos es esencial para cualquier formato por simple que sea. El resultado más importante del encuentro australiano fue el triunfo dialéctico de los estructuralistas ya que se definieron los calificadores: subelemento (SUBELEMENT), lengua (LANG), y el calificador esquema (SCHEME), con carácter de "necesarios" aunque siempre opcionales, que se usarán para refinar la semántica del conjunto de elementos del DC de modo que faciliten una descripción más precisa. A partir del encuentro de Canberra, se diferenció entre un modelo calificado (*qualified DC*) y un modelo más simple, que no incluía el uso de los calificadores subelemento, esquema (*scheme*) y lengua y que se denominó, por ello, *unqualified DC* (Cfr. *infr.* Ejemplo1. DC)¹²². Por otra parte, fue también en Canberra donde se planteó el problema del multilingüismo en Internet, al respecto de este tema se acordó desarrollar variedades del DC en otros idiomas distintos al inglés.

- **DC-5** (octubre 1997. Helsinki, Finlandia). OCLC/*National Library of Finland*): A este taller del Dublin Core, se le denomina, según la literatura sobre este tema, con la aliteración en inglés *the Finnish finish* (el final finlandés), ya que la semántica del DC se dio por finalizada en Helsinki. Este encuentro constituyó, por otra parte, la base para la primera estandarización formal del DC, que afianzó

¹²¹ Cfr. Capítulo 3, Nota 119.

¹²² Las dos corrientes, minimalista y estructuralista, han dado lugar a múltiples estudios y posiciones en torno a la aplicación del modelo Dublin Core. Deacon, por ejemplo, alude a que la inconsistencia y parcialidad en la aplicación de este formato de metainformación se debe al conflicto de ambas posturas (minimalista y estructuralista) y apela a una mayor simplicidad en la descripción porque un alto grado de estructura será demasiado costoso si no se realiza a la par de las facilidades de recuperación. Deacon basa su estudio en la comparación del modelo de catalogación con el modelo de las bases de datos de citas. Prue Deacon. *Simplicity vs. Structure: Which Way for the Dublin Core?* *Cataloguing Australia*, 1999, vol. 25, nº 1/4, p. 36-43.

el camino hacia una mayor implementación de este formato. Los aspectos más importantes que se acordaron allí, fueron: una codificación útil para la información cronológica mediante la constitución del elemento fecha (DC.Date) según la norma ISO 8601; un acuerdo mayor en la utilización del elemento *cobertura* (DC.Coverage) para soportar la búsqueda por referencia espacial o geográfica o para identificar un periodo de tiempo, pudiendo asociarse a ese campo un calificador de esquema (SCHEME) en el futuro. En cuanto al elemento relación (DC.Relation), se dictó en Helsinki, "el principio 1:1" que abordaba el problema de asignación de metadatos coherentes a los recursos que derivan de otros y que tienen también sus propios metadatos, resolviendo asignar un conjunto discreto de metadatos para cada objeto usando la sintaxis del HTML y, sobre todo, apoyándose en la confianza, que ya desde entonces existía, en las capacidades del RDF como formato apto para la viabilidad de ese principio en el futuro.

- **DC-6** (Noviembre 1998, Washington): El sexto *workshop* del DC profundizó en algunos temas que ya se habían estado barajando en sesiones anteriores. En Washington se gestaron objetivos que, desde nuestro punto de vista, han culminado posteriormente. En primer lugar, la formalización del DC. Se acordó crear una Agencia para el mantenimiento de esta iniciativa, tratando de emular el funcionamiento del W3C o del IETF, para conseguir un funcionamiento estable y consolidar la cultura de consenso internacional del DC¹²³. Así, se crearía el DC Directorate, entonces albergado en la Oficina de OCLC <<http://purl.org/dc>>, un *Policy Advisory Committee* (PAC) y un *Technical Advisory Committee* (TAC). Se planteó asimismo, el tema de la estandarización formal, iniciada a través del RFC

¹²³ Este es por ejemplo uno de los objetivos que, a pesar de esbozarse en el 98, han culminado formalmente en la creación del sitio DCMI <<http://dublincore.org>>, accesible en la Web, según muestran sus metadatos <<http://dublincore.org/index.shtml.rdf>>, desde el 16 de enero de 2001.

2413 del IETF¹²⁴, que se dirigiría tanto a la NISO como al CEN para que el conjunto de elementos de metadatos (DCMES) fuesen estándares futuros, tanto a nivel americano como europeo. Se instó también la formulación de una especificación para codificar la semántica del DC en HTML¹²⁵. Se estudiaron las implicaciones de la cualificación de los metadatos DC para la interoperabilidad a través de la utilización de *schemes* o vocabularios controlados que permitieran aumentar la especificidad semántica. Igualmente, comenzaron los estudios de las relaciones entre el DC y otras iniciativas de metadatos (*crosswalks*). Finalmente, se discutieron los posibles cambios en la estructura del DC, abogando por una mayor coherencia en la expresión (por ejemplo que bajo los elementos del DC existiese un modelo lógico como el expresado en los *Functional Requirements for Bibliographic Records* (FRBR) de la IFLA) y se pensó en una segunda versión del formato (*Dublin Core 2.0*).

- **DC-7** (Octubre 1999, Frankfurt, Alemania)/ **DC-8** (Octubre 2000, Ottawa, Canadá): Podríamos decir que en estos dos talleres, así como en la actividad desarrollada en torno al Dublin Core a lo largo de esos dos años, se marca la estabilidad de dicho modelo de metadatos. El aspecto que caracterizó los desarrollos del Dublin Core en el bienio 1999-2000 es la aprobación del Consorcio Web, en febrero de 1999, del modelo y la sintaxis de RDF como recomendación. En julio de ese mismo año el DCMES 1.1 se convirtió en una

¹²⁴ La primera expresión formal de la semántica del DC es el RFC 2413 donde se describe el DC 1.0 es decir, la semántica de los 15 elementos del DC. Stuart Weibel, et al. *Dublin Core Metadata for Resource Discovery: RFC 2413* [documento TXT]. IETF, Network Working Group, September 1998. Disponible en: <http://www.ietf.org/rfc/rfc2413.txt> (consultado el 11 de agosto de 2000).

¹²⁵ Lo que culminaría en diciembre de 1999 en el RFC 2731. *Op. cit.*, <http://www.ietf.org/rfc/rfc2731.txt>, pero se basaría en otros trabajos previos de, entre otros, Weibel. Stuart Weibel. *A proposed convention for embedding metadata in HTML* [documento HTML]. En: *Distributed indexing/searching workshop (1996. Cambridge)*. World Wide Web Consortium, 2 de junio de 1996, rev. 20 de junio de 1996. Disponible en: <http://www.w3.org/Search/9605-Indexing-Workshop/ReportOutcomes/S6Group2.html>; también disponible en: <http://staff.oclc.org/~weibel/html-meta.html> (consultados el 18 de marzo de 2001).

recomendación del DCMI¹²⁶. En esta etapa sedimentaron algunas de las líneas iniciadas en encuentros anteriores, como la consolidación de la normalización formal, tanto a través de la votación del, entonces, proyecto de norma Z39.85 en el marco de la ANSI/NISO, como del acuerdo del CEN (CWA 13874), lo que constituirá el camino lógico hacia una normalización ISO del DCMES. Desde el *workshop* de Frankfurt se creó una discusión en torno a los calificadores que culminaría en julio de 2000 en un documento donde se recogen todas las propuestas de calificación de los elementos¹²⁷, bien a través de la precisión de los mismos (p. ej., el elemento `relation` podría tener las siguientes matizaciones: `IsVersionOf`, `HasVersión`, `IsReplacedBy`, `Replaces`, `IsRequiredBy`, `Requires`, `IsPartOf`, `HasPart`, `IsReferencedBy`, `References`, `IsFormatOf`, `HasFormat`), bien a través del establecimiento de esquemas para el contenido de los elementos (así, p. ej., para el elemento `subject`, se podrían utilizar los siguientes *schemes*, LCSH, MeSH, DDC, LCC, UDC). También en estos años, importantes iniciativas en el marco de los metadatos adoptaron el modelo DC (como es el caso de la OAI, *Open Archives Initiative*) o lo tuvieron en cuenta para sus correspondencias e interoperabilidad (como el MPEG7).

En el **DC-8** (Ottawa, 2000), destaca la explicación del DC a partir de la gramática¹²⁸ y, sobre todo, el concepto de *application profile* introducido por Rachel Heery¹²⁹ como un tipo de esquema (*schema*) de metainformación basado

¹²⁶ Dublin Core Metadata Element Set, Version 1.1: Reference Description (DCMI Recommendation, 2 de julio de 1999) <<http://dublincore.org/documents/1999/07/02/dces>>.

¹²⁷ Dublin Core Qualifiers (DCMI Recommendation, 11 de julio de 2000) <<http://dublincore.org/documents/2000/07/11/dcmes-qualifiers>>.

¹²⁸ Thomas Baker. A Grammar of Dublin Core [documento HTML]. *D-Lib Magazine*, October 2000, vol. 6, nº 10. Disponible en: <http://www.dlib.org/dlib/october00/baker/10baker.html> (consultado el 3 de enero de 2001).

¹²⁹ La presentación PPT de Heery en el DC-8 se puede consultar en: <http://dublincore.org/workshops/dc8/pp/dc8-ap-v3.ppt> Una explicación más exhaustiva del concepto de los perfiles de aplicación para la gestión de esquemas de metadatos se publicó en la revista electrónica Ariadne: [cont.]

en el DC. A raíz de la introducción de este concepto, se han desarrollado múltiples "perfiles de aplicación", como por ejemplo el de las bibliotecas (DC-Lib¹³⁰), de forma que sirvan de intercambio en los contextos informativos que utilizan distintos formatos y/o estándares de metadatos. La definición de perfiles de aplicación, no es más que una forma de seguir el modelo conceptual de Warwick y establecer una arquitectura de metadatos que soporte la visión modular de esta tecnología, consolidada por la actividad del W3C en torno a RDF.

- Finalmente, en octubre del 2001 se celebró la *International Conference on Dublin Core and Metadata Applications* en Tokio (**DC-9, DCMI-2001**). A pesar de que aún no se hayan publicado las conclusiones, el título de la conferencia y los contenidos de las ponencias presentadas¹³¹, son evocadores de la legitimación del concepto de *application profile* y del reconocimiento de los distintos perfiles que se pueden/deben realizar a tenor de este modelo de metainformación. En este encuentro del DCMI se discutió sobre la creación de perfiles de aplicación así como mecanismos para expresarlos en RDF. Esta nueva concepción del DC podría llevarnos a pensar en él como un modelo de propósito general, tal y como

Rachel Heery and Manjula Patel. Application Profiles: Managing Metadata Schemas [documento HTML]. *Ariadne*, 9-October-2000, issue 25. Disponible en: <http://www.ariadne.ac.uk/issue25/app-profiles> (consultado el 5 de julio de 2001). *Vid.* Glosario, *application profile*.

¹³⁰ DC-Lib (DCMI Working Draft, 8 de agosto de 2001): <http://www.dublincore.org/documents/2001/08/08/library-application-profile> (consultado el 10 de agosto de 2001); DC-Lib (DCMI Working Draft, 12 de octubre de 2001): <http://www.dublincore.org/documents/2001/10/12/library-application-profile> (consultado el 10 de noviembre de 2001). Este perfil de aplicación, uno de los más consolidados en la comunidad de la DCMI, supone uno de los primeros logros del Grupo de Trabajo sobre bibliotecas, creado en el seno de la iniciativa DCMI en 1999 en el *workshop* de Frankfurt.

¹³¹ *DC-2001 Proceedings of the International Conference on Dublin Core and Metadata Applications* [documento PDF]. Keizo Oyama and Hironobu Gotoda, eds. Tokyo: National Institute of Informatics, 2001. Disponible en: <http://www.nii.ac.jp/dc2001/proceedings/product/book.pdf> (consultado el 12 de noviembre de 2001).

lo clasificamos en esta tesis, pero con distintas proyecciones de aplicación específica¹³².

El hecho de que esta investigación haya ido creciendo a la par que el Dublin Core, nos permite establecer dos fases en el desarrollo de los metadatos, que se corresponden, a su vez, con las dos fases que advertimos en el desarrollo del modelo DC. Por su popularidad o/y aceptación, este modelo ha marcado el desarrollo de los metadatos y por ende, también define dos momentos en el desarrollo y planteamiento de los objetivos de esta tesis. La evolución de los metadatos se puede resumir pues, a tenor del DC, en estas dos etapas:

- 1) La constitución de estándares *de facto*¹³³ para optimizar la recuperación de la información en la Red (1995-1998).
- 2) La normalización formal y la preocupación por la interoperabilidad y el acceso global a la información (1999-2001). En esta etapa se tratará de soportar la amplia diversidad de semánticas de metadatos en un marco estructural y sintáctico común. A la consolidación de este planteamiento contribuirá enormemente que, en 1999 el RDF se convirtiera una recomendación del Consorcio Web.

¹³² Este tema ya había sido objeto de discusión un año antes en el ámbito del proyecto SCHEMAS de la Unión Europea, en el taller celebrado en Bonn en Noviembre de 2000. Manjula Patel. *SCHEMAS: Second Workshop Report: Publishing and Sharing Your Metadata Application Profile* [documento PDF]. Schemas Forum, 26 de enero de 2001. Disponible en: <http://www.schemas-forum.org/workshops/ws2/wks2.pdf> (consultado el 4 de abril de 2001).

SCHEMAS es un forum surgido en el seno del V Programa Marco y del programa IST (*Information Society Technologies*) que tiene como finalidad ayudar a los posibles implementadores a entender y utilizar los estándares de metainformación de tal forma que sean compatibles con los principios de "buen uso" de metadatos. Tendremos oportunidad de hablar más profusamente de este proyecto en el capítulo siguiente cuando analicemos la formalización de normas de metadatos. (*Vid.* 6.2.) De todas formas, hay que señalar que SCHEMAS trabaja en cooperación con la DCMI para construir un registro para la gestión de *namespaces*. *Vid.* Glosario, *namespace*.

¹³³ *Vid.* la discusión en torno al nivel de formalización de los modelos de metadatos en el apartado 6.2. de esta tesis y, más concretamente, el proceso de normalización formal del DCMES, en el apartado 6.3.

Los formatos de metadatos son ese tipo de cosas que se explican y comprenden más fácilmente a través de ejemplos. Por ello, a continuación expondremos un ejemplo¹³⁴ del DC de la "primera etapa", codificado en una sintaxis HTML. Se combinan pues, la sintaxis general de HTML a través de la etiqueta <META> y un modelo semántico, *a priori*, de propósito general. El ejemplo consta de dos partes: en primer lugar se presenta una descripción simple (*Unqualified DC*), válida en cualquier versión del HTML a partir de la 2.0; en segundo lugar, se presenta una descripción más compleja de este mismo recurso Web, aumentada con los calificadores de Canberra (*Qualified DC*), que incluye la versión extendida del DC que se menciona desde la especificación del HTML 4.0¹³⁵. En ambos casos, el conjunto de metadatos asociados a dicho recurso Web, puede estar almacenado en la propia estructura del HTML en la sección de la cabecera de un documento Web <HEAD> . . . </HEAD>:

Ejemplo DC en HTML (1)

➤ *Unqualified Dublin Core:*

```
<META NAME= DC.Creator CONTENT="Rebecca Bradshaw">
<META NAME=DC.Title CONTENT="ADAM: Art, Design, Architecture and
Media Information Gateway">
<META NAME=DC.Publisher CONTENT="University of Northumbria at
Newcastle on behalf of the ADAM Consortium">
<META NAME=DC.Subject CONTENT="ADAM, Art, design, architecture,
media, eLib, JISC">
<META NAME=DC.Description CONTENT= "ADAM: an information gateway to
quality-assured resources on the Internet in art, design,
architecture and media. An Access to Network Resources project
funded as part of JISC' s Electronic Libraries Programme.">
<META NAME=DC.Date CONTENT="1997-05-22">
<META NAME=DC.Identifier CONTENT="http://adam.ac.uk/">
<META NAME=DC.Language CONTENT="en">
<META NAME=DC.Format CONTENT="text/html">
<META NAME=DC.Rights CONTENT="http://adam.ac.uk/adam/rights.html">
```

¹³⁴ Ejemplo tomado de: Tony Gill. The Dublin Core Metadata Element Set: Useful Tool or the Emperor's Newest Clothes? *Spectra*, Winter 1997, p. 45.

¹³⁵ World Wide Web Consortium. *HTML 4.0 Specification... Op. cit.*, <http://www.w3.org/TR/1998/REC-html40-19980424>

➤ Qualified Dublin Core:

```
<META NAME=DC.Creator CONTENT="Rebecca Bradshaw">
<META NAME=DC.Creator.Email CONTENT="rebecca@adam.ac.uk">
<META NAME=DC.Title LANG="en" CONTENT="ADAM: Art, Design,
Architecture and Media Information Gateway">
<META NAME=DC.Publisher LANG="en" CONTENT=="University of
Northumbria at Newcastle on behalf of the ADAM Consortium">
<META NAME=DC.Subject LANG="en" CONTENT="ADAM, Art, design,
architecture, media, eLib, JISC">
<META NAME=DC.Description LANG="en" CONTENT="ADAM: an information
gateway to quality-assured resources on the Internet in art, design,
architecture and media. An Access to Network Resources project
funded as part of JISC' s Electronic Libraries Programme.">
<META NAME=DC.Date.LastModified SCHEME="ISO 31-1" CONTENT="1997-05-
22">
<META NAME=DC.Identifier CONTENT="http://adam.ac.uk/">
<META NAME=DC.Language SCHEME="ISO 639" CONTENT="en">
<META NAME=DC.Format SCHEME="IMT" CONTENT="text/html">
<META NAME=DC.Rights CONTENT="http://adam.ac.uk/adam/rights.html">
```

A continuación, presentamos otro ejemplo asignando el conjunto de metadatos DC con calificadores¹³⁶, que describiría esta tesis si se tratase de un documento HTML que hubiésemos publicado en la Web, en la dirección imaginaria que utilizamos sobre este mismo ejemplo en el capítulo anterior <<http://www.bib.uc3m.es/~mendez/tesis/tesis.htm>>. La forma de codificación o almacenamiento de la descripción del ejemplo es, al igual que en el caso anterior, HTML, ya que queremos resaltar la combinación de estructuras de información de carácter general (HTML) con el DC como modelo de metadatos de propósito general. Utilizaremos otras formas de representación de la semántica del DC en RDF/XML al hablar del *Resource Description Framework* como "metamodelo" de metadatos¹³⁷.

¹³⁶ La forma de codificar DCq en documentos HTML es curiosamente todavía (10/11/01) un "borrador de trabajo", mientras que la codificación del DCq en XML/RDF es una propuesta de recomendación, lo que demuestra que hay un mayor interés en el seno de la DCMI por regularizar el uso del DC en la nueva y flexible sintaxis XML/RDF que en la obsoleta HTML. *Cfr.*

- Recording Qualified Dublin Core Metadata in HTML Meta Elements (DCMI Working Draft, 15 de agosto de 2000): <http://www.dublincore.org/documents/2000/08/15/dcq-html>
- Expressing Qualified Dublin Core in RDF / XML (DCMI Proposed Recommendation, 29 de agosto de 2001): <http://www.dublincore.org/documents/2001/08/29/dcq-rdf-xml>

¹³⁷ *Vid.* 5.2.3. *Cfr.* Ejemplo RDF (2): DC en sintaxis RDF.

Ejemplo DC en HTML (2)

```

<html>
<head profile="http://dublincore.org/qdcmes/1.0">
<meta name="DC.Title" lang="es" content="metadatos y recuperación de
información: estándares, problemas y aplicabilidad en bibliotecas
digitales">
<meta name="DC.Creator.Personalname" lang="es"
content="Eva María Méndez Rodríguez">
<meta name="DC.Creator.Personalname.address" lang="es"
content="emendez@bib.uc3m.es">
<meta name="DC.Subject" lang="es"
content="metadatos, recuperación de información, bibliotecas
digitales, estándares, normas, Biblioteca Virtual Cervantes,
sistemas de información de calidad, estructuración de la
información, metainformación">
<meta name="DC.Description" lang="es"
content="Tesis doctoral donde se trata de mostrar la utilidad de los
metadatos para la recuperación de información, sobre todo en
contextos finitos de información o bibliotecas digitales que
permiten diseñar sistemas de información de calidad. Se estudian
distintos formatos, normas y problemas relacionados con este tema.
además se plantea un proyecto de aplicación de metadatos a la
Biblioteca Virtual Miguel de Cervantes.">
<meta name="DC.Publisher.Corporatename" lang="es"
content="Universidad carlos III de Madrid. Departamento de
Biblioteconomía y Documentación">
<meta name="DC.Contributor.personalname" lang="es" content="José A.
Moreiro González">
<meta name="DC.Date" scheme="ISO8601" lang="es"
content="2001-12-13">
<meta name="DC.Type" lang="es" content="tesis">
<meta name="DC.Format" scheme="IMT" lang="es" content="text/html">
<meta name="DC.Identifier" scheme="URI"
content="http://www.bib.uc3m.es/~mendez/tesis/tesis.htm">
<meta name="DC.Language" lang="es" scheme:"RFC1766" content="es-ES">
<meta name="DC.Relation.Isreferencedby" lang="es"
content="http://www.bib.uc3m.es/~mendez/profesional/index.htm">
<meta name="DC.Relation.Requires" scheme="URI"
content="http://www.bib.uc3m.es/~mendez/tesis/tesis.css">
<meta name="DC.Relation.isFormatOf" scheme="URI"
content="http://www.bib.uc3m.es/~mendez/tesis/tesis.doc">
<meta name="DC.rights" lang="es" content="Eva Méndez">
</head>
</html>

```

A tenor de estos ejemplos, podemos decir que el formato del DC carece de complejidad de creación, interpretación y almacenamiento. Como expresamos anteriormente, está diseñado originalmente para que el autor y/o el editor de un recurso pueda generar los metadatos que describan su propia información; tarea que se ve favorecida por las múltiples herramientas que facilitan la asignación y edición de metadatos, como *Reggie Metadata Editor* o *Dublin Core Metadata Template*

(*Nordic Metadata Project*) que hemos seleccionado y analizado en el capítulo anterior.

El DC presenta un control bibliográfico a medio camino entre el nivel de detalle del formato MARC o el estructurado modelo TEI —del que hablaremos a continuación— y las metaetiquetas básicas de HTML que utilizan algunos servicios de indización automática de la WWW, basándose en los valores `keywords` y `description` del atributo `name`. Por ello, su caracterización responde al nivel medio o zona 2 del *continuum* señalado por Dempsey y Heery¹³⁸. Se incluye en la banda 2 porque se adapta a la caracterización genérica de tal zona al *no presentar estructuras internas muy elaboradas y contener el nivel de designación suficiente para la búsqueda por algunos campos*. Sin embargo, debemos matizar que, a pesar de que el Dublin Core se aplica a objetos discretos al igual que el resto de modelos de esta zona, difiere de otros formatos incluidos en ella en que sí recoge las relaciones existentes entre ellos a través del elemento de relación (`DC.Relation`¹³⁹).

La ventaja del DC frente al resto de formatos reside, amén de su facilidad de uso, en que es un modelo internacionalmente aceptado y es el primer candidato para la descripción de recursos interdisciplinares, de modo que tiende a convertirse en un modelo semántico de datos a través de RDF. Otro de los avances que supone este esquema tiene que ver con los protocolos de recuperación que utiliza. En 1998, por ejemplo, el ZIG¹⁴⁰ (*Z39.50 Implementers Group*) acordó insertar los 15 elementos del DC en la lista de atributos de uso del protocolo Z39.50 Bib-1. Esto implicaba que

¹³⁸ Vid. 3.5. y Tabla 8.

¹³⁹ Cfr. *supr.*, Tabla 12. El elemento `Relation` fue expresamente definido en el DC-5, donde se constituyó además el grupo de trabajo para desarrollar la lista de elementos a la que ya nos hemos referido, para seleccionar las relaciones y así asegurar la interoperabilidad. Dublin Core Qualifiers. *Op. cit.*, <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers>

¹⁴⁰ Ralph Le Van. *Dublin Core and Z39.50: Draft Version 1.2* [documento HTML]. OCLC, Office and Research Special Projects, 2 de febrero de 1998. Disponible en: <http://cypress.dev.oclc.org:12345/~rrl/docs/dublincoreandz3950.html> (consultado el 9 de julio de 1998).

las distintas versiones de los clientes Z39.50 podrían especificar las búsquedas a los elementos del DC. Las soluciones que se desarrollaron en este sentido, estaban dirigidas a posibilitar la recuperación basada en los elementos del DC en tres sintaxis de registros Z39.50: USMARC, HTML y GRS-1. Esta propuesta está esperando el consenso de la comunidad del Dublin Core¹⁴¹ en lo que respecta a los calificadores y esquemas, y al consenso del ZIG en la revisión de la norma Z39.50-2001 para la nueva arquitectura de atributos¹⁴². Sin embargo, se están desarrollando distintas interfaces que permiten la búsqueda basada en el protocolo Z39.50 en registros DC codificados en XML¹⁴³.

Para concluir esta breve aproximación a la DCMI, debemos mencionar que son muchos los proyectos que legitiman este modelo como estándar para la descripción de recursos electrónicos y que demuestran que es un esquema de metadatos válido para describir DLOs de distintas áreas del conocimiento, principalmente: arte y humanidades, bibliografía, educación, medio ambiente, matemáticas, medicina, ciencia y tecnología, e incluso comercio electrónico¹⁴⁴. Asimismo, se han creado

¹⁴¹ El trabajo del ZIG en virtud de la aplicación del DC está considerado en la actualidad una Nota del DCMI. Ralph Le Van. *Dublin Core and Z39.50* [documento XHTML]. DCMI, 2 de febrero de 1998. Disponible en: <http://www.dublincore.org/documents/1998/02/02/dc-z3950> (consultado el 10 de noviembre de 2001).

¹⁴² ANSI/NISO Z39.50-2001. *Information Retrieval (Z39.50): Application Service Definition and Protocol Specification* [documento PDF]. Bethesda, Maryland: NISO, rev. 21 de noviembre de 2001. Disponible en: <http://www.loc.gov/z3950/agency/revision/part1.pdf> y <http://www.loc.gov/z3950/agency/revision/part2.pdf> (consultado el 8 de diciembre de 2001).

¹⁴³ Tal es el caso, por ejemplo, de ZETOC <<http://zetoc.mimas.ac.uk>>, un servicio que da acceso Z39.50 a los sumarios electrónicos de la base de datos de artículos de la British Library que consigue la interoperabilidad entre el DC y otros modelos de codificación de registros bibliográficos, basándose en una sintaxis XML y en el Dublin Core como esquema de metadatos básico. Para más información sobre este prototipo, Vid. Ann Apps and Ross MacIntyre. *zetoc: a Dublin Core Based Current Awareness Service* [documento PDF]. *Journal of Digital Information*, 2001, vol. 2, issue 2. Disponible en: <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Apps/apps-v2.pdf> (consultado el 15 de noviembre de 2001).

¹⁴⁴ Para más información sobre los proyectos que usan el Dublin Core como su estándar de metadatos. Vid. Dublin Core Projects: <http://dublincore.org/projects>

distintos grupos de trabajo para implementar diversos perfiles de aplicación del DCMI a las diversas áreas temáticas de la Red, de entre los cuales merece ser destacado, teniendo en cuenta el enfoque de esta tesis, el grupo dedicado a la implementación del DC en bibliotecas¹⁴⁵. Desde el punto de vista de las aplicaciones informáticas para la creación de metadatos, como ya quedó patente en el capítulo anterior, el DC es, también uno de los formatos más aceptados. Estos niveles de aplicación y aceptación de la DCMI como esquema de metainformación en bibliotecas digitales y servicios de información aumentarán a partir de ahora a través del influjo que ejercerá el hecho de ser una norma formal (Z39-85).

5.2.2. Metadatos de propósito específico

A pesar de que el Dublin Core puede adaptarse para describir cualquier tipo de recurso electrónico, las diversas comunidades de usuarios de la Red y las distintas áreas de aplicación requieren diferentes elementos y niveles de complejidad. Por este motivo, son muchas las instituciones y organismos que están desarrollando esquemas de metadatos complementarios, basados en el DC en algunos casos (p. ej., EdNa, AGLS), o totalmente distintos (como la TEI Header).

De todas formas, esos formatos *específicos de un dominio*¹⁴⁶, pretenden ajustarse a la temática y particularidades de la información electrónica para la que se formulan. En este apartado vamos a tratar un ejemplo paradigmático de este tipo de modelos: la *Text Encoding Initiative* (TEI). También mencionaremos otros esquemas

¹⁴⁵ Guinchard presentó en abril de 2001 un interesante estudio basado en un cuestionario enviado a distintas bibliotecas de todo el mundo para analizar el nivel de aplicación del DC, según diversos puntos de vista. Carolyn Guinchard. Summary of DC-Libraries Questionnaire Responses [documento RTF]. En: *Survey Results: Dublin Core Use in Libraries* [correo electrónico en lista de distribución]. *DC-LIBRARIES*, Miércoles 25 de abril de 2001; 16:03:45. Disponible en: <http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind0104&L=dc-general&F=&S=&P=2403> (consultado el 22 de mayo de 2001).

¹⁴⁶ Lo que nosotros hemos denominado "metadatos de propósito específico" es llamado por algunos autores "*domain-specific metadata*". V.gr. Virpul Kashyap, Amit Sheth. *Information Brokering... Op. cit.*, p. 20; Sherry L. Vellucci. *Metadata. Op. cit.*, p. 204.

de metadatos, algunos con una estructura sintáctica prescrita y otros diseñados independientemente de la sintaxis que utilicen como un modelo semántico adaptable a varias estructuras de información, principalmente y de forma creciente, XML y RDF.

5.2.2.1. TEI Header (Text Encoding Initiative)

La Iniciativa de Codificación Textual (TEI) es algo más que un mero formato de metadatos banda-3 basado en SGML, como hemos comentado hasta ahora. Inicialmente fue un proyecto internacional de investigación incoado en 1987, patrocinado originalmente por tres asociaciones profesionales de amplio prestigio internacional: la ACH (*Association for Computers and Humanities*), la ALLC (*Association for Computational Linguistics*) y la ACL (*Association for Literary and Linguistic Computing*) y consolidado por la Dotación Nacional Americana para las Humanidades (NEH) y la DG XIII de la Comisión Europea a través del III Programa Marco Europeo para la Investigación Lingüística e Ingeniería. En un principio fue un proyecto de investigación, pero hoy en día se ha convertido en un Consorcio internacional (TEI-C, 2000).

La iniciativa TEI supuso un esfuerzo internacional conjunto para elaborar una DTD de SGML adaptada al ámbito académico y humanístico. Su objetivo principal, declarado por Burnard¹⁴⁷ en diferentes contribuciones sobre dicho formato, consiste en definir un conjunto de directrices generales para la representación de materiales textuales en formato electrónico, de tal forma que permita a los investigadores de cualquier disciplina intercambiar y reutilizar los recursos, independientemente del software y del hardware que utilicen y de su campo de aplicación. De forma más específica, Burnard y Light, matizan el objetivo inicial de TEI diciendo que este proyecto trata de: *Definir un conjunto de recomendaciones para la codificación de*

¹⁴⁷ Lou Burnard es el editor europeo de TEI desde 1990 y es uno de los desarrolladores de la especificación desde ese año. *Vid.* Bibliografía.

*materiales textuales literarios y lingüísticos en formato electrónico para estandarizar el trabajo existente y facilitar el progreso adecuado en un campo en desarrollo vertiginoso*¹⁴⁸.

De 1987 a 1994, un grupo formado por numerosos investigadores de distintas nacionalidades trabajó en el proyecto TEI con un doble objetivo: acordar y codificar la representación digital de textos de diversa índole y, al mismo tiempo, definir un mecanismo que fuera extensible para abarcar toda la gama de prácticas de codificación en el ámbito académico. El resultado fue la recomendación TEI (*TEI Guidelines*) de mayo de 1994 que, según el propio Burnard¹⁴⁹, dichas recomendaciones no sólo se publicaron y aceptaron por la comunidad científica, sino que también significaron uno de los más completos análisis emprendidos en el ámbito de las estructuras textuales.

Tal y como se publicaron en 1994, las *TEI Guidelines* estaban formadas por 1300 páginas que definían y documentaban aproximadamente 600 elementos de SGML que podían combinarse de múltiples formas para crear distintas DTDs. Por ello, en 1998 se hizo una revisión de ese copioso conjunto de etiquetas para formar una DTD más operativa y modular, lo que se conoce como TEI P3¹⁵⁰ —donde la "P" significa "para el Público"—, que permite manejar la mayoría de los textos a un nivel de detalle razonable. No obstante, se reconoció que debían personalizarse distintos

¹⁴⁸...to define a set of recommendations for the encoding of literary and linguistic textual materials in electronic form, both in order to standardize existing work, and to facilitate the development of good practice in a rapidly developing field. Lou Burnard and Richard Light. *Op. cit.*, <http://www.ukoln.ac.uk/metadata/BIBLINK/wp1/sgml>

¹⁴⁹ Lou Burnard. Text Encoding for Interchange: a New Consortium [documento HTML]. *Ariadne*, 23-Jun-2000, issue 24. Disponible en: <http://www.ariadne.ac.uk/issue24/tei> (consultado el 8 de agosto de 2000).

¹⁵⁰ *TEI Guidelines for Electronic Text Encoding and Interchange (P3)* [documento HTML]. Virginia: University of Virginia, Electronic Text Center, [1998?]. Disponible en: <http://etext.virginia.edu/TEI.html> (consultado el 19 de agosto de 2000). No obstante, esta versión "más completa" de TEI comenzó su revisión y adaptación a XML en mayo de 1999. Vid. <http://www.tei-c.org/P4beta/index.htm> (consultado el 19 de agosto de 2000).

perfiles para comunidades particulares de usuarios. Una de esas versiones personalizadas de TEI es TEI-U5, más conocida como TEI Lite¹⁵¹, que se desarrolló específicamente para dirigir las necesidades del conjunto más importante de usuarios de TEI, las bibliotecas digitales o los centros de textos electrónicos, sin que ello obste para que la Iniciativa de Codificación Textual haya tenido otros usos en el mercado de documentos electrónicos¹⁵². A pesar de todas estas reducciones de la DTD original de TEI, es preciso comentar que, en la práctica, cada biblioteca digital que selecciona este modelo de descripción de documentos, adopta su propia DTD basándose en las necesidades de la documentación que codifica¹⁵³, normalmente con la ayuda de la aplicación *The Pizza Chef: a TEI Tag Set Selector*¹⁵⁴. Sin embargo, la más profunda y definitiva revisión de la DTD, es la adaptación de TEI P3 a XML (TEI P4)¹⁵⁵ en la que se ha trabajado a raíz de la formación, en diciembre de 2000,

¹⁵¹ *TEI U5: Encoding for Interchange: An Introduction to the TEI* [documento HTML]. TEI Consortium, 1995, rev. 21 de enero de 2001. Disponible en: <http://www.hcu.ox.ac.uk/TEI/Lite> (consultado el 7 de julio de 2001).

¹⁵² Por ejemplo, el *Center for Electronic Text in the Law* (CETL) <<http://www.law.uc.edu/CETL>> es un proyecto del College of Law de la Universidad de Cincinnati que utiliza la DTD de SGML (TEI P3) para codificar documentos de archivo relativos a los derechos humanos.

¹⁵³ Así, por ejemplo, la Biblioteca Virtual *Miguel de Cervantes* <http://cervantesvirtual.com> ha adoptado su propia DTD en XML de TEI denominada *cervante.dtd*. Cfr. 9.1.2.

¹⁵⁴ The Pizza Chef: a TEI Tag Set Selector: <http://www.hcu.ox.ac.uk/TEI/pizza.html>

¹⁵⁵ Como hemos explicado en la primera parte de este capítulo, XML simplifica algunas de las características de SGML. Por ello y dado que TEI incluía algunas de esas características, fue fácil adaptar este esquema a XML, de la misma forma que ha ocurrido con otros formatos también descritos originalmente en SGML, como por ejemplo EAD o CIMI. Para seguir el estado de la revisión de TEI P4, que tiene previsto completarse a finales de 2001, Vid. <http://www.tei-c.org/P4X/Status>

Recientemente, en el Congreso Computing Arts 2001, celebrado en la Universidad de Sydney del 26 al 28 de Septiembre de 2001, Lou Burnard presentó una ponencia donde destacó la especial aptitud de TEI para adecuarse a la flexibilidad del nuevo XML, y del trabajo del TEI-C para adaptar la especificación a XML (P4). Se refirió a los desarrollos futuros de TEI en este contexto, como un nuevo entorno para la construcción de vocabularios XML compatibles para distintas áreas de investigación. Lou Burnard. *TEI and XML: A Marriage Made in Heaven. An introduction to The Future of Digital Information* [documento HTML]. En: *Computing Arts: Digital Resources for Research in the Humanities* (2001. Sydney), University of Sydney, Research Institute for Humanities [cont.]

del Consorcio TEI (TEI-C) que, desde entonces, se encarga de desarrollar este estándar de marcado.

A pesar de sus orígenes de propósito lingüístico y eminentemente volcado a la información textual, las recomendaciones TEI servirán para la estructuración de la información electrónica de todo tipo, incluidas imágenes. Uno de los componentes más significativos de TEI —al menos el que más nos interesa en el contexto de esta investigación—, es el denominado **cabecera TEI** (*TEI Header*, TEIH), que permite definir una descripción bibliográfica detallada para cada texto codificado TEI como parte de su codificación original, que puede ser utilizada por los investigadores para el análisis textual o como control bibliográfico en el caso de que exista un gran número de archivos de texto.

Sin embargo, la mayoría de los autores, al describir y tratar la cabecera TEI desde una perspectiva documental o bibliográfica —Shieh¹⁵⁶, Burnard y Light¹⁵⁷, por ejemplo— inciden en la idea de que la cabecera TEI es una información para los catalogadores. Se refieren a ella como *página de título electrónica*, argumentando que la descripción del texto y su codificación facilitan un análogo electrónico a la página de título (portada) en un trabajo impreso o un equivalente al código del libro o manual introductorio, que acompaña al conjunto de datos electrónicos. La cabecera TEI puede utilizarse para construir registros de bases de datos bibliográficas relativos a textos codificados TEI, o como *cabecera independiente* para describir recursos de la Red que no estén codificados en TEI. Esta última utilidad (cabecera independiente)

& Social Sciences, octubre 2001. Disponible en: <http://www.arts.usyd.edu.au/Arts/departs/rihss/keyabs.html#Burnard> (consultado el 13 de noviembre de 2001).

¹⁵⁶ Jackie Shieh. The TEI Header and the Cataloging Rules [documento HTML]. En: American Library Association. *Committee on Cataloging: Description and Access Task Force on Metadata and Cataloging Rules: Final Report*. ALA, 3 de junio de 1998, rev. 4 de junio de 1998. Disponible en: <http://www.ala.org/alcts/organization/ccs/ccda/tf-tei4.htm> (consultado el 18 de junio de 1998).

¹⁵⁷ Lou Burnard and Richard Light. *Op. cit.*, <http://www.ukoln.ac.uk/metadata/BIBLINK/wp1/sgml>

ha convertido al TEI en un formato más de metadatos con capacidad para representar el contenido estructurado de cualquier documento, incluidos los recursos de Internet.

El grado de dificultad para crear esta cabecera TEI depende fundamentalmente de la cantidad de información que se incluya en ella y del nivel de normalización del contenido según reglas externas como AACR2. Como señalan Dempsey y Heery¹⁵⁸ en su estudio general sobre los formatos de descripción de recursos, *si una cabecera independiente se crea de tal forma que tenga el mismo contenido que un registro MARC, con la misma adherencia a la práctica catalográfica, requerirá el mismo nivel de destreza exigido para la catalogación bibliotecaria*. En cualquier caso, podemos decir que la cabecera TEI, sin ser un formato especialmente abstruso, requiere cierto nivel de conocimiento del modelo, amén de la mayor o menor normalización bibliotecaria que se establezca. La complejidad de aplicación se incrementa porque no existen herramientas de software específico para asignar metadatos conforme al modelo TEIH¹⁵⁹.

Por otra parte, el nivel de detalle de la cabecera TEI es bastante flexible y es posible crear registros de metadatos¹⁶⁰ no conformes con las AACR2 o con cualquier otra versión nacional de las ISBD. De hecho, así se reconoce en las propias recomendaciones TEI. Independientemente del nivel de detalle posible que se elija, del grado de normalización que se estime necesaria utilizando reglas bibliográficas, o de que, por el contrario, se realice una descripción en texto libre, es importante

¹⁵⁸ Lorcan Dempsey and Rachel Heery. *A Review of Metadata...* *Op. cit.*, <http://www.ukoln.ac.uk/metadata/desire/overview/overview.rtf> Esta afirmación, en realidad, podría trasladarse a cualquier formato de metadatos. Para evaluar la cabecera TEI como fuente de catalogación de recursos basada en las AACR2. *Vid.* Jackie Shieh. *Op. cit.*, <http://www.ala.org/alcts/organization/ccs/ccda/tf-tei4.htm>

¹⁵⁹ Por el contrario, sí existe software específico para crear, modificar, etc. documentos según TEI — *Vid. v. gr.* <http://xtalk.opensource.ru:8101/SGML/TEItools/index-en.html>— que puede usarse para tratar también cabeceras TEIH, amén de que pueden utilizarse herramientas genéricas de edición y validación de documentos XML (p. ej., XMetal que trata XML desde una perspectiva orientada a documentos, más que a datos).

¹⁶⁰ *Vid.* Glosario, *metadata record*.

definir un nivel unívoco ya que, en entornos distribuidos de información como Internet, la utilidad de la codificación TEI se vería afectada por problemas de interoperabilidad a la hora de compartir registros.

Como cabría esperar de un formato basado en la norma SGML y adaptado a XML, la cabecera TEI está formada por un conjunto de elementos y atributos agrupados en conjuntos de etiquetas. Un elemento (en este caso) puede ser una unidad, como el título o el autor, mientras que el atributo da información sobre una ocurrencia particular de ese elemento, pudiendo establecerse como par atributo/valor. Así, por ejemplo, en la descripción del perfil existe un elemento `<textClass>` que identifica los encabezamientos de materia, y el vocabulario controlado usado se describe por un atributo `<keywords scheme= LSCH>`. Este modelo de metadatos, se estructura en cuatro partes:

- 1) Descripción del fichero (`<FILEDESC> . . . </FILEDESC>`), que incluye las características bibliográficas del texto codificado así como de la fuente¹⁶¹ (`<SOURCEDESC> . . . </SOURCEDESC>`). Esta parte de la cabecera TEI es la que más se aproxima a la descripción bibliográfica tradicional. No en vano contiene información estructurada detallada trazada en las normas y en la práctica tradicional de catalogación y está modelada sobre los estándares de catalogación bibliotecaria¹⁶². Consta de tres elementos obligatorios (mención de título, mención de publicación y fuente del texto electrónico) y de cuatro opcionales (mención de edición, extensión, mención de serie, y notas).
- 2) Descripción de la codificación (`<ENCODINGDESC> . . . </ENCODINGDESC>`) donde se explican los niveles de codificación y análisis del texto y la relación

¹⁶¹ Recordemos que la cabecera TEI es una parte de una DTD de SGML o XML que puede tener un uso independiente para la descripción bibliográfica de documentos electrónicos, pero su origen es la codificación de textos literarios o de ciencias humanas, con lo cual siempre habrá un texto fuente al que es preciso hacer referencia. Cuando se utiliza como cabecera independiente tiene la misma estructura pero más dirigida a la representación de contenido.

¹⁶² Vid. Jackie Shieh. *Op. cit.*, <http://www.ala.org/alcts/organization/ccs/ccda/tf-tei4.htm>

entre el fichero electrónico y la fuente en que se basa (incluye decisiones editoriales sobre cómo se codifica el texto y detalles sobre el proceso editorial, como el tratamiento de las líneas en blanco, tabulaciones, etc.).

- 3) Descripción del perfil (<PROFILEDESC>...</PROFILEDESC>). En esta parte de la cabecera se incluye información adicional no bibliográfica sobre el contexto en el cual se ha producido el documento textual (lenguaje, detalles de los participantes, clasificación de materias, etc.).
- 4) Descripción de la revisión (historia de la descripción¹⁶³) (<REVISIONDESC>...</REVISIONDESC>), donde se incluirán detalles de actualizaciones, correcciones del texto, etc., que permiten al codificador revelar la historia de los cambios producidos durante el desarrollo del texto electrónico.

Estas partes principales de la codificación TEIH están compuestas, a su vez, por distintos elementos como se muestra en la tabla a continuación (Tabla 13)¹⁶⁴:

ELEMENTO - TEIH	DESCRIPCIÓN y USO
<TEI Header>	Comienzo de la cabecera TEI que puede llevar diversos atributos, como el tipo de norma utilizado, por ejemplo: <TEI Header type="AACR2">, el autor de la cabecera TEI <TEI Header autor="Eva Méndez"> o la fecha de creación de TEIH.
<fileDesc> (descripción del fichero digital/fuente)	
<title type=_____>	Título principal o uniforme: <titleType="uniform">o <titleType="main">.
<author>	Autor de la fuente original en la forma establecida

¹⁶³ Lou Burnard and Richard Light. *Op. cit.*, <http://www.ukoln.ac.uk/metadata/BIBLINK/wp1/sgml>

¹⁶⁴ La tabla 13 muestra algunos de estos elementos. Para confeccionarla nos hemos basado en el capítulo 5 de las *TEI Guidelines* en XML: <http://www.tei-c.org/P4X/HD.html> y en la DTD completa de la TEIH en XML: <http://www.tei-c.org/P4X/DTD/teihdr2.dtd> Nótese como se hace una explicación de los elementos desde una perspectiva eminentemente bibliotecaria en aras a reflejar tácitamente la comparación que se hace de este modelo de cabecera de metadatos con el estándar MARC. No obstante, según las distintas DTDs adaptadas a cada biblioteca digital, estos elementos pueden variar, tanto en su declaración, como en su uso posterior en el marcado de documentos específicos de una colección. *Cfr.* 9.3.2.

ELEMENTO - TEIH	DESCRIPCIÓN y USO
	(siguiendo las RRCC: Apellidos, Nombre).
<code><editor></code>	Editor de la fuente original (electrónica o impresa).
<code><respStmt></code>	Otras menciones de responsabilidad (editor, compilador, ilustrador).
<code><editionStmt></code>	Mención de edición. Se refiere a la edición electrónica, no a la edición original (impresa).
<code><extent></code>	Extensión del fichero electrónico del DLO (en la forma: "ca.nkb")
<code><publicationStmt></code>	Área de edición del archivo electrónico. (Al ser un elemento Stmt incluirá otros propios de las ISBD ¹⁶⁵)
<code><publisher></code>	En TEI, se entiende editor a cualquiera que haya recopilado el texto electrónico y tomado decisiones sobre él. El editor de la versión digital de un DLO puede ser personal o corporativo.
<code><distributor></code>	El distribuidor es quien hace accesible el texto electrónico, esto es, el personal técnico que hace efectiva la publicación electrónica.
<code><idno></code>	Identificador numérico de la publicación determinado por el editor del documento digital.
<code><availability></code>	Disponibilidad. Se utilizarán elementos especiales cuando se anticipe que el encabezamiento TEI se va a compartir en usos locales.
<code><date></code>	Fecha de publicación del documento electrónico. Normalmente en formato AAAA o AAAAMMDD
<code><seriesStmt></code>	Si es posible, se establece el control autorizado para series creadas localmente.
<code><notesStmt></code>	Elemento opcional que dependerá de las decisiones sobre visualización. Suele utilizarse también para indicar que elementos como el título, autor, etc., se han atribuido.
<code><sourceDesc></code>	Este elemento marca el inicio de la descripción del documento fuente original en que basamos la edición

¹⁶⁵ Los elementos que terminan por "Stmt" (`<editionStmt>`, `<titleStmt>`) normalmente encierran un grupo de elementos especializados que contienen información estructurada sobre ellos; por ejemplo en el caso de los elementos bibliográficos, se utiliza esa terminación (Stmt) con los nombres de elementos que corresponden a las áreas de las ISBD.

ELEMENTO - TEIH	DESCRIPCIÓN y USO
	electrónica.
<bibl> <biblStruct> <biblFull>	Se presenta como elemento preferido <biblFull> para permitir la búsqueda en partes de la descripción de la fuente.
<biblFull> <title>	Título de la fuente. Es un elemento repetible que permite describir títulos alternativos.
<biblFull> <author>	Autor de la obra fuente. Si el nombre del autor/es difiere de la forma establecida para mencionar al autor, se debe incluir la forma <author type="alternate">.
<biblFull> <editionStmt>	Mención de edición tal y como aparece en la fuente original digitalizada.
<biblFull> <extent>	Descripción física del documento fuente en que se basa el DLO que estamos describiendo.
<biblFull> <publicationStmt> <publisher>	Editor del documento original. Elemento no repetible. Si la fuente original es una coedición, se harán constar los distintos editores separados por ";"
<biblFull> <pubPlace>	Lugar de publicación del documento fuente. Elemento no repetible. Igual que el editor, si son varios los lugares de publicación se harán constar bajo la misma etiqueta separados por ";"
<biblFull> <date>	Fecha que consta en el documento original. Normalmente el año en formato AAAA.
<biblFull> <idno>	Número de identificación de la fuente original. Normalmente, ISBN, etc. aunque también puede utilizarse para indicar la localización de la fuente en una colección tradicional de una institución ¹⁶⁶ .
<biblFull>	Declaración de la serie; en su caso, del documento

¹⁶⁶ El elemento <idno> dentro de la etiqueta <biblFull> de la descripción de la fuente (<sourcedesc>), puede utilizarse alternativamente para indicar que el documento original sobre el que se está basando la versión electrónica, procede de un fondo particular. Tendrá un número de identificación en ese fondo, que se hará constar si éste tiene una forma normalizada, v. gr., <idno type="LC call number">. Este sería el caso, por ejemplo, de la BVC, en la que proyectamos la aplicación de nuestra tesis: los documentos del portal de Argentina, prestados por la BN de ese país a efectos de digitalización, tendrán un número de identificación dentro de ese fondo. Cfr. Capítulo 9.

ELEMENTO - TEIH	DESCRIPCIÓN y USO
<code><seriesStmt></code>	original.
<code><biblFull></code> <code><notes></code>	En este elemento se harán constar notas de cualquier índole, pero siempre relativas al documento original o documento fuente.
<code><encodingDesc></code> (descripción de la codificación)	
<code><projectDesc></code>	En este elemento se hará constar una descripción del proyecto, es decir, la finalidad por la cual se ha codificado el fichero electrónico.
<code><editorialDecl></code>	En el elemento relativo a la "declaración editorial", como evoca el acrónimo de la etiqueta que lo describe, se incluirá una explicación sobre cómo se ha tratado el fichero electrónico, a fin de hacer constar decisiones técnicas de edición electrónica.
<code><refsDecl></code>	Este es el espacio que concede la TEIH a los metadatos administrativos ¹⁶⁷ (p. ej., mantenimiento).
<code><classDecl></code> <code><taxonomy id=____></code>	En su caso, este elemento permite identificar la clasificación que se utiliza. El <code>id=</code> , tiene una función semejante al atributo <code>scheme</code> en la codificación HTML del DC. Así, por ejemplo: <code><taxonomy id=LCSH></code> .
<code><profileDesc></code> (descripción del perfil)	
<code><creation></code> <code><date></code>	Fecha tal como la concibe el creador del documento digital.
<code><langUsage></code>	Idioma que utiliza el agente creador del documento digital en la forma normalizada de denominación de idiomas. (p. ej. eng-US)
<code><language id=____></code>	Identificación de la norma utilizada para codificar el idioma (ISO 639-2, que también se emplea para la codificación MARC en el campo 041).
<code><textClass></code>	Números de la clasificación del texto digital.
<code><keywords></code>	En el caso de que no se adopte ningún lenguaje controlado o clasificación, se usará esta etiqueta para

¹⁶⁷ En el capítulo 3 definimos los metadatos administrativos como aquellos utilizados para la gestión y administración de los recursos digitales en el momento de su creación o introducción en el sistema, útiles para comprobar el mantenimiento de recursos Web, como por ejemplo, el control de las versiones de un documento, etc. *Cfr.* 3.5.

ELEMENTO - TEIH	DESCRIPCIÓN y USO
<code><term></code>	consignar las palabras claves asociadas al documento.
<code><keywords scheme=____> <term></code>	Calificación de la indización según un vocabulario controlado específico o clasificación de materias, en consonancia con el identificador de la clasificación utilizado en el elemento <code><encodingDesc></code> .
<code><revisionDesc></code> (revisión de la descripción)	
<code><change> <resp> <item></code>	Etiquetas que sirven para controlar los cambios y revisiones en la descripción de metadatos. Se preferirá el uso de códigos para definir el estado de la revisión más que expresiones textuales en lenguaje natural.

Tabla 13. Interpretación de los elementos principales de la TEIH en XML

Para crear este tipo de cabecera, bien sea asociada a un documento conformado TEI, bien relativa a cualquier otro documento electrónico (cabecera independiente), además de las Recomendaciones TEI (Lite, P3 o P4), se hace alusión a las AACR2 y las ISBD para completar el registro formal, cuando la información se incluye en los elementos apropiados; sin embargo, a la propia flexibilidad de TEI, se añade la flexibilidad implícita en XML, permite incluir información en texto libre.

No obstante, como analizaremos a continuación, cada biblioteca o colección de textos marcados en TEI (SGML/XML) utiliza un nivel de codificación en la cabecera de metadatos¹⁶⁸ (tanto desde el punto de vista de los elementos descriptivos de la TEIH que adopta, como de las normas, reglas, etc. en que se basa para completar dichos elementos). Así, el registro de metainformación conforme al TEIH no responde a los mismos elementos en los dos ejemplos siguientes:

- 1) Texto electrónico de la obra *Liberty Lyrics* (1895) de Louisa Sarah Bevington, realizado con la recomendación de la cabecera TEI en SGML (P3)

¹⁶⁸ Este concepto *nivel de codificación* al que aludimos en el texto, está íntimamente relacionado con el concepto de *granularity* que discutimos en el capítulo anterior. Cfr. 4.1, Vid. Glosario.

dentro del proyecto *Victorian Women Writers*¹⁶⁹ del *Library Electronic Text Resource Service* (LETRS) de la Universidad de Indianápolis (Ejemplo TEI Header 1).

- 2) Registro de metadatos del texto digital *Las esmeraldas*, de Joaquín Dicenta, en la Biblioteca Virtual *Miguel de Cervantes*¹⁷⁰, realizado en función a la codificación TEI en XML (P4) (Ejemplo TEI Header 2).

Ejemplo TEI Header (1)

```
<TEI HEADER>
<FILEDESC>
  <TITLESTMT>
    <TITLE>Liberty Lyrics (1895): a machine-readable
    transcription</TITLE>
    <AUTHOR>Bevington, Louisa Sarah (Guggenberger) (1845-
    ?)</AUTHOR> <RESPSTMT>
    <RESP>Transcribed and encoded by </RESP> <NAME>Felix
    Jung</NAME>
    </RESPSTMT>
    <RESPSTMT><RESP>Edited by </RESP>
    <NAME>Perry Willett</NAME></RESPSTMT>
  </TITLESTMT>
  <EXTENT>TEI formatted filesize uncompressed&colon; 1426
  bytes</EXTENT>
  <PUBLICATIONSTMT>
  <PUBLISHER>Library Electronic Text Resource Service (LETRS),
  Indiana University </PUBLISHER>
  <DATE>September 22, 1995</DATE>
  <AVAILABILITY>
  <P>&copy; 1995, The Trustees of Indiana University. Indiana
  University makes a claim of copyright only to original
  contributions made by the Victorian Women Writers Project
  participants and other members of the university community.
  Indiana University makes no claim of copyright to the original
  text. Permission is granted to download, transmit or otherwise
  reproduce, distribute or display the contributions to this
  work claimed by Indiana University for non&hyphen;profit
  educational purposes, provided that this header is included in
  its entirety. For inquiries about commercial uses, please
  contact&colon;
  <ADDRESS>
  <ADDRLINE>Library Electronic Text Resource Service</ADDRLINE>
  <ADDRLINE>Main Library</ADDRLINE>
  <ADDRLINE>Indiana University</ADDRLINE>
  <ADDRLINE>Bloomington, IN 47405</ADDRLINE>
  <ADDRLINE>United States of America</ADDRLINE>
```

¹⁶⁹ Victorian Women Writers Project: <http://www.indiana.edu/~letrs/vwwp/index.html>

¹⁷⁰ Biblioteca Virtual Miguel de Cervantes (BVC): <http://www.cervantesvirtual.com> Vid. Capítulo 9.

```

<ADDRLINE>Email: LETRS@indiana.edu</ADDRLINE>
</ADDRESS>
</P>
</AVAILABILITY>
</PUBLICATIONSTMT>
<SERIESSTMT>
<TITLE>Victorian Women Writers Project&colon; an Electronic
Collection</TITLE>
<RESPSTMT><NAME>Perry Willett, </NAME>
<RESP>General Editor</RESP></RESPSTMT>
</SERIESSTMT>
<SOURCEDESC>
<BIBLFULL>
<TITLESTMT>
<TITLE>Liberty Lyrics </TITLE>
<RESPSTMT><RESP>by </RESP>
<NAME>L.S. Bevington</NAME></RESPSTMT></TITLESTMT>
<EXTENT>16 p.</EXTENT>
<PUBLICATIONSTMT>
<PUBLISHER>Printed and Published by James Tochatti,
</PUBLISHER>
<PUBLISHER>&ldquo;Liberty&rdquo; Press </PUBLISHER>
<PUBPLACE>London </PUBPLACE>
<DATE>1895</DATE>
</PUBLICATIONSTMT>
</BIBLFULL>
<P>The copy transcribed is from Michigan State University
Libraries.</P>
</SOURCEDESC>

```

</FILEDESC>

<ENCODINGDESC>

```

<EDITORIALDECL>
<P>All poems occur as DIV0. Sonnets are attributed as
"type=sonnets"; the rest are "type=poem". All quotation marks,
hyphens, dashes, apostrophes and colons have been transcribed
as entity references. All < lg > (line groups) are attributed
as cantos, stanzas, couplets, verse paragraphs, etc. All poems
with regularly indented lines use the attribute "rend" in the
< l > tag, with the value "indent1" for one tab stop,
"indent2" for two tab stops, etc. All split lines are
attributed as "type=i" for the initial portion, and "type=f"
for the final portion.</P>
<P>All apostrophes and single right quotation marks are
encoded as &rsquo;.</P>
<P>Any hyphens occurring in line breaks have been removed; all
hyphens are encoded as &hyphen; and em dashes as &mdash;.</P>
</EDITORIALDECL>
<TAGSDECL>
<TAGUSAGE GI="back" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="body" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="corr" OCCURS="4"></TAGUSAGE>
<TAGUSAGE GI="div" OCCURS="3"></TAGUSAGE>
<TAGUSAGE GI="div0" OCCURS="15"></TAGUSAGE>
<TAGUSAGE GI="div1" OCCURS="2"></TAGUSAGE>
<TAGUSAGE GI="docauthor" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="docdate" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="docimprint" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="doctitle" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="emph" OCCURS="15"></TAGUSAGE>
<TAGUSAGE GI="front" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="head" OCCURS="18"></TAGUSAGE>
<TAGUSAGE GI="L" OCCURS="484"></TAGUSAGE>

```

```

<TAGUSAGE GI="lg" OCCURS="109"></TAGUSAGE>
<TAGUSAGE GI="p" OCCURS="7"></TAGUSAGE>
<TAGUSAGE GI="pb" OCCURS="14"></TAGUSAGE>
<TAGUSAGE GI="text" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="titlepage" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="titlepart" OCCURS="1"></TAGUSAGE>
<TAGUSAGE GI="titlestmt" OCCURS="2"></TAGUSAGE>
</TAGSDECL>
</ENCODINGDESC>

```

[Nota: Este ejemplo no incluye descripción del perfil]

```

<REVISIONDESC>
  <CHANGE>
    <DATE>1995-06-30</DATE>
    <RESPSTMT><NAME>Felix Jung, </NAME>
    <RESP>editor.</RESP></RESPSTMT>
    <ITEM>finished data entry, basic encoding and
    proofing</ITEM></CHANGE>
  <CHANGE>
    <DATE>1995-09-11</DATE>
    <RESPSTMT><NAME>Perry Willett, </NAME>
    <RESP>general editor.</RESP></RESPSTMT>
    <ITEM>finished TEI-conformant encoding and final
    proofing</ITEM></CHANGE>
</REVISIONDESC>
</TEI HEADER>

```

En este ejemplo, la cabecera TEI (TEIH) forma parte de todo el texto de la obra *Liberty Lyrics*, también marcado según las Recomendaciones TEI. Se puede apreciar cómo se combinan en el mismo marco descriptivo la información catalográfica tradicional, información sobre los derechos y el acceso, detalles específicos de codificación y la versión de la información que se presenta. El ejemplo no incluye la descripción del perfil (lenguaje, fecha, etc.) probablemente porque en el marco del proyecto concreto (*Victorian Women Writers Project*), estos datos no aportarían demasiada información.

La versión final para la Web, en HTML, de este documento¹⁷¹, no contempla sin embargo ninguna metainformación (el único elemento de la cabecera HTML que aparece es: <TITLE>Victorian Women Writers Project</TITLE>).

¹⁷¹ Vid. <http://www.letrs.indiana.edu/cgi-bin/vwwp-query.pl?type=bibl&rgn=TEXT&idno=lnU-AJE6190>

Ejemplo TEI Header (2)

```

<?xml version="1.0"?>
<?xml-stylesheet href="http://sprod/intranet/xmltei/cervante.xsl"
type="text/xsl"?>
<!DOCTYPE TEI.2 SYSTEM "http://sprod/intranet/xmltei/cervante.dtd"
[
<!ENTITY % ISolat1 SYSTEM "http://sprod/intranet/xmltei/iso-
lat1.ent"> %ISolat1;
<!ENTITY % ISolat2 SYSTEM "http://sprod/intranet/xmltei/iso-
lat2.ent"> %ISolat2;
<!ENTITY % ISOpub SYSTEM "http://sprod/intranet/xmltei/iso-pub.ent">
%ISOpub;
<!ENTITY % ISOnum SYSTEM "http://sprod/intranet/xmltei/iso-num.ent">
%ISOnum; ]>
<TEI.2>
<TEI Header day.created="20" month.created="7" year.created="1999">
<fileDesc>
  <titleStmt>
    <title type="main">Las esmeraldas</title>
    <author>Joaquín Dicenta</author>
  </titleStmt>
  <publicationStmt>
    <publisher>Biblioteca Virtual Miguel de Cervantes
Saavedra</publisher>
    <pubPlace>Universidad de Alicante</pubPlace>
    <idno>001179</idno>
    <availability status="free">
      <p>Copyright &copy; Universidad de Alicante, Banco
Santander Central Hispano 1999-2001. Accesible desde
http://cervantesvirtual.com</p>
    </availability>
    <eDate day="20" month="7" year="1999"/>
  </publicationStmt>
  <sourceDesc>
    <p></p>
  </sourceDesc>
</fileDesc>
<profileDesc>
  <langUsage>
    <language id="es">Español</language>
  </langUsage>
</profileDesc>
<revisionDesc>
  <change>
    <eDate day="24" month="4" year="2001"/>
    <respStmt><name>Carmen Hernández Segovia</name>
    <resp>Correctora</resp>
  </respStmt>
  <item>Marcado del texto</item>
</change>
</revisionDesc>
</TEI Header>

```

Igual que ocurría en el ejemplo anterior, en la versión final para la Web, las metaetiquetas que aparecen en el documento HTML no corresponden a la estructura

de metadatos de la TEI Header. El código fuente del documento del ejemplo, tal y como se visualiza en la WWW, es el siguiente¹⁷²:

```
<TITLE>Las esmeraldas - Biblioteca Virtual Miguel de Cervantes</TITLE>
<META NAME="Title" CONTENT="Las esmeraldas - Dicenta, Joaquín">
<META NAME="Author" CONTENT="Biblioteca Virtual Miguel de Cervantes">
<META NAME="Owner" CONTENT="webmaster@cervantesvirtual.com">
<META NAME="Revisit" CONTENT="60 days">
<META NAME="Keywords" CONTENT="biblioteca, virtual, Cervantes, literatura, gratis, Las esmeraldas, Dicenta, Joaquín">
<META NAME="Description" CONTENT="Edición electrónica incluida en la Biblioteca Virtual Miguel de Cervantes, con un fondo de 3000 obras de literatura clásica Española e Hispanoamericana, de acceso gratuito">
<META NAME="Robots" CONTENT="nofollow">
<META NAME="Language" CONTENT="Spanish">
```

La no-correspondencia de la cabecera TEI con el resultado del código fuente creado (a partir del marcado TEI) para la Web en HTML, corrobora varias de las teorías y posturas que hemos adoptado en esta tesis:

- Los metadatos, en el ámbito de las bibliotecas digitales implicadas en el proceso de digitalización y creación del documento¹⁷³, se asignan en la etapa de pre-publicación de los documentos, siendo éstos un conjunto finito de DLOs para los que se ha proyectado un marcado específico sobre los datos y el texto.
- El objetivo del uso de metainformación en bibliotecas virtuales o en colecciones digitales de textos no responde a la finalidad, normalmente entendida, del uso de metaetiquetas HTML para mejorar la visibilidad en los motores de búsqueda genéricos de Internet.
- Por otra parte, también justificamos con este ejemplo, la decisión de no considerar y valorar en nuestra investigación el nivel de utilización de metadatos, ya que es imposible analizar los metadatos internos del sistema (creados según un

¹⁷² Vid. <http://cervantesvirtual.com/servlet/SirveObras/059871994744848133373113>

¹⁷³ Vid. 8.2.1.

formato particular) a través de la información que suministra el código fuente final visualizado en el navegador.

Finalmente, en términos de recuperación basada en las estructuras de la TEIH, es importante comentar que el software para la visualización y recuperación en esquemas de codificación basados en SGML es escaso¹⁷⁴ y tampoco existen protocolos de búsqueda y recuperación de información que expresamente operen sobre las TEIH independientes. En relación con este tema, se ha discutido el uso del protocolo Z39.50 para la búsqueda y recuperación de información sobre cabeceras TEI ya que, al igual que el resto de los formatos banda-3, su perfil es adecuado para el acceso a las colecciones digitales. La reformulación de TEI P3 en XML (P4), además de simplificar la publicación de documentos en la Web a través de XSLT y de herramientas de publicación como Cocoon¹⁷⁵, permite augurar instancias de recuperación genérica en la Web también para este formato.

A pesar de que la adaptación de la DTD de TEI a XML es relativamente reciente, son muchos los proyectos que utilizan este lenguaje de descripción de documentos o de marcado para codificar las representaciones electrónicas de documentación textual siguiendo en algunos casos todavía, la TEI P3 de SGML, y la TEIH para describir la metainformación asociada. TEI se ha convertido en un estándar *de facto* en muchos de los proyectos de digitalización de textos literarios, lingüísticos y, en general, de contenido humanístico¹⁷⁶, permitiendo dotar a esos documentos electrónicos de un marcado semántico muy detallado. De igual modo, la

¹⁷⁴ De 1995 a 1998 se desarrollaron algunas herramientas para la visualización de SGML en la WWW como Panorama de Softquad. Sin embargo, la complejidad estructural del metalenguaje SGML ha hecho que las empresas de software concentren sus esfuerzos en desarrollar aplicaciones para XML.

¹⁷⁵ Cocoon es un entorno flexible de publicación de XML que se basa en la separación del contenido, la lógica y el estilo de documentos XML. Vid. <http://xml.apache.org/cocoon>

¹⁷⁶ El Consorcio TEI <<http://www.tei-c.org>> tiene registrados 64 proyectos en 31 idiomas distintos, relativos a 11 áreas de las Humanidades y de las ciencias sociales (literatura, historia, música, corpus lingüísticos, etc.). Para aproximación a cada uno de ellos Vid. <http://www.tei-c.org/Applications>

conformación de estos conjuntos de textos en bibliotecas digitales se ha basado, en muchos casos, en la cabecera TEI que incluye la DTD (tanto de la versión P3, como de la P4).

Desde nuestro punto de vista, una de las razones que pueden justificar la adopción de la TEIH para fundar la organización de documentos en las bibliotecas, es el alto componente bibliográfico asociado a este modelo de metadatos y las facilidades de interoperabilidad de esta codificación y el formato bibliotecario por excelencia (MARC). La alta proyección de la cabecera TEI como modelo de metadatos para describir recursos electrónicos, sobre todo en los primeros años del DC (1995-98), puede deberse al nivel de detalle que permite este esquema desde el punto de vista estrictamente catalográfico, lo que llevó a muchos sectores del mundo bibliotecario a equipararlo al formato MARC¹⁷⁷.

5.2.2.2. Otros esquemas de metadatos

Además de la capacidad de TEI para expresar metainformación a través de la TEIH en textos humanísticos, existen múltiples soluciones de metadatos, adaptados a la sintaxis XML¹⁷⁸, que recogen las necesidades de descripción de la diversidad de los recursos electrónicos en la Red o de una comunidad de usuarios particular. Esta

¹⁷⁷ Esta visión de la TEIH como un "formato de catalogación" más que como un modelo de metadatos (entendidos éstos en el sentido más estricto que defendemos en esta tesis) se puede apreciar en los trabajos del *Task Force on Metadata and the Cataloging Rules* en el seno de la ALA o en contribuciones como la de Pouchard, que habla de la compatibilidad entre las TEIH y el USMARC basándose en el nivel de detalle (*granularity*) de las cabeceras TEI y la flexibilidad del MARC. Line Pouchard. *Cataloging for Digital Libraries: The TEI Scheme and the TEI Header* [documento HTML]. *Katharine Sharp Review*, Winter 1998, issue 6. Disponible en: <http://mirrored.ukoln.ac.uk/lis-journals/review/review/6> (consultado el 9 de julio de 1999).

¹⁷⁸ Para profundizar en cada uno de estos formatos, y en otros muchos de propósito más específico, existe una amplia bibliografía actualizada de cada uno de ellos en Internet: descripciones, evaluaciones, proyectos concretos, etc. *Vid.* Bibliografía e índice de siglas y acrónimos, donde se pueden encontrar, junto al significado de la sigla correspondiente, el URL principal de cada modelo o propuesta de metainformación.

multiplicidad y variedad de formatos se debe, en parte, a que existen, como hemos visto, varias aproximaciones técnicas a la metainformación y, sobre, todo a que cada disciplina o cada tipo de información necesita metadatos específicos. Así, se están desarrollando paralelamente distintas soluciones, esquemas o modelos de metainformación con la misma finalidad de hacer útiles sus datos, pero con motivaciones y necesidades diferentes. Citamos a continuación algunos de estos modelos que sirvan para ilustrar las dos razones que hemos apuntado para la proliferación de estándares de propósito específico: a) formatos de la información electrónica diversos, y b) especificidad temática de la información que albergan los DLOs, todo ello unido a la propia inercia en la gestión de la información electrónica, abocada a la pluralidad de esquemas de metainformación.

a) Modelos de metadatos específicos para un tipo/formato de información

En este grupo de esquemas debemos destacar, sobre todo, los dirigidos a representar la imagen digital, bien sea fija o en movimiento, y los destinados a describir la información sonora y/o multimedia.

Los objetos multimedia de la Red tienen múltiples características que justifican la necesidad de contar con un modelo de metadatos que refleje su complejidad y su multidimensionalidad. Con este objetivo surgen estándares como el *Multimedia Content Description Interface*, más conocido como **MPEG-7**, que ha sido desarrollado por el *Moving Picture Experts Group*. MPEG-7, al igual que su antecesor MPEG-4, permite un nivel de descripción de metadatos con un gran potencial para organizar colecciones digitales de audio y vídeo. Se trata de un estándar que describe unidades (DLOs) de audio, vídeo o audio y vídeo, que ha normalizado un conjunto de descripciones y un lenguaje particular para especificarlas, denominado DDL (*Description Definition Language*). Existen otras

iniciativas para la descripción de contenidos multimedia como por ejemplo el XML orientado a Objetos (**SOX**)¹⁷⁹.

Para el no menos complejo mundo de la imagen, también existen diversas iniciativas y modelos de metadatos que tratan de contemplar las características de la imagen digital o digitalizada. En este ámbito destaca la actividad iniciada en abril de 1999 en el seno del DIG (*Digital Imaging Group*) para crear un mecanismo normalizado que permita al usuario final ver el uso de la imagen digital tan fácil y tan flexible como los métodos fotográficos tradicionales, facilitando al mismo tiempo los beneficios que sólo son posibles con un formato digital¹⁸⁰. Con ese mismo espíritu se planteó, en agosto de 2000, la primera versión del modelo **DIG35**¹⁸¹, que trata de aprovechar la flexibilidad de XML para describir la imagen digital a través de un conjunto público de metadatos divididos en cinco grupos lógicos: parámetros básicos de la imagen, creación, metadatos descriptivos del contenido, históricos (relativos al proceso de producción de la imagen digital) y propiedad intelectual. Paralelamente, en julio de ese mismo año, la **NISO** había publicado un documento de

¹⁷⁹ En el 8º Congreso del W3C, celebrado en Canadá en 1999, encontramos una excelente revisión de la descripción de recursos multimedia, así como una propuesta de adaptación del DC y el MPEG7 a RDF, realizadas por dos investigadoras australianas. Vid. Jane Hunter, Liz Armstrong. A Comparison of Schemas for Video Metadata Representation [documento HTML]. En: *International World Wide Web Conference (8. 1999. Toronto)*, rev. 5 de diciembre de 2000. Disponible en: <http://www8.org/w8-papers/3c-hypermedia-video/comparison/comparison.html> (consultado el 18 de marzo de 2001).

¹⁸⁰ Digital Imaging Group. *The Power of Metadata Is Propelling Digital Imaging Beyond the Limitations of Conventional Photography* [documento PDF]. DIG, Agosto de 1999. Disponible en: <http://members.digitalimaging.org/sheredocs/DIG35WhitePaper.pdf> (consultado el 8 de agosto de 2000).

¹⁸¹ Digital Imaging Group. *DIG35 Specification: Metadata for Digital Images. Version 1.0* [documento PDF]. DIG, 30 de agosto de 2000. Disponible en: <http://members.digitalimaging.org/shareddocs/downloads/dig35v1.0-sept00.pdf> (consultado el 18 de mayo de 2001).

trabajo para definir el diccionario de datos de metadatos técnicos aplicados a la imagen¹⁸².

Un formato que podríamos incluir tanto en este grupo como en el siguiente, dedicado a los modelos de metadatos específicos por el área temática a la que se aplican, es el esquema **VRA Core Categories**, ya que es aplicable tanto a las obras de arte como a las imágenes que las representan. Desarrollado por la Asociación de Recursos Visuales (VRA) y basado en las CDWA (*Categories for the Description of Works of Art*), contempla un conjunto sencillo de categorías para describir, gestionar y proporcionar un acceso rápido a la información del objeto de arte original y a su imagen visual o trasunto documental codificado, sea éste una fotografía o una transparencia¹⁸³.

¹⁸² NISO Draft Standard. *Data Dictionary: Technical Metadata for Digital Still Images* [documento PDF]. NISO, 5 de julio de 2000. Disponible en: <http://www.niso.org/pdfs/DataDict.pdf> (consultado el 30 de mayo de 2001). Este documento de trabajo de la NISO recoge una lista de elementos de datos relevantes en las tareas vinculadas al ciclo de vida de la imagen digital desde el punto de vista exclusivamente técnico (tono, color, tamaño, etc.). Es un subconjunto del formato DIG35 que aspira a convertirse en un estándar formal NISO para la descripción y el intercambio de imágenes digitales.

¹⁸³ Sobre VRA, *Vid. The Core Categories for Visual Resources: Version 2.0, October 15, 1997* [documento HTML]. Visual Resources Association, Data Standards Committee, rev. 20 de octubre de 1997. Disponible en: <http://www.oberlin.edu/~art/vra/guide.html> (consultado el 24 de agosto de 1998). Visual Resources Association Data Standards Committee. *VRA Core Categories: Version 3.0* [documento HTML]. Harvard University, 24 de julio de 2000. Disponible en: <http://www.gsd.harvard.edu/~staffaw3/vra/vracore3.html> (consultado el 8 de agosto de 2000).

Sobre los estándares de metadatos aplicables a los objetos de arte, realizamos un estudio previo a este trabajo haciendo hincapié en las normas para el acceso e intercambio de objetos de información en el ámbito de los museos digitales. En dicho estudio se analizan, a parte del VRA, otros metadatos para las artes visuales (como Object ID y CDWA) y esquemas específicos para la información museística y de patrimonio cultural (CIMI, SPECTRUM). Se realiza asimismo, una comparación de las categorías principales de los modelos CDWA y VRA Core 2.0. *Vid. Eva M^a Méndez, Murtha Baca. Acceso al patrimonio cultural en el mundo digital: estándares para la recuperación de información en museos virtuales. Revista de Museología (Dossier: Museos del siglo XXI), n^o 21, 1^{er} cuatrimestre 2001, p. 68-77.*

b) Modelos de metadatos específicos para un área temática

Podríamos decir, sin que fuese demasiado arriesgado, que prácticamente cualquier colectivo de usuarios de Internet vinculados a un área temática —CVU (Comunidad Virtual de Usuarios) o CoI (*Community of Interest*)— ha desarrollado su propio modelo de metadatos o ha adaptado uno de carácter general, como el Dublin Core, a las particularidades de la información que manejan o describen. Sin embargo, a modo de ejemplo, vamos a citar esquemas de metainformación de las tres áreas temáticas, desde nuestro punto de vista, más afectadas por la gestión de información electrónica y, por ello, con un nivel mayor de desarrollo de metadatos: la educación, la información gubernamental y la información geoespacial¹⁸⁴.

El efecto de Internet en la educación ha provocado muchos cambios en los modelos de aprendizaje, así como en la forma de proporcionar acceso a los materiales curriculares. Se han creado por ello, múltiples estándares (*schemas* y *schemes*) para mejorar el potencial de la educación basada en Internet. Uno de esos modelos es **IMS** (*Instructional Management Systems*) surgido de Educom y de la *National Learning Infrastructure Initiative* (NLII). La estructura de metadatos del IMS consta de categorías, elementos de datos y tipos de datos, contenidos en una estructura jerárquica y codificados en XML¹⁸⁵. Otros esquemas de metadatos aplicados a la información educativa, son: **EdNA** (*Education Network Australia*) basado en el

¹⁸⁴ De hecho, a estas tres áreas se ha dedicado una sesión monográfica en el último congreso del DC celebrado en Tokyo. *DC-2001 Proceedings... Op. cit.*, <http://www.nii.ac.jp/dc2001/proceedings/product/book.pdf> (Special Session 1: Metadata in Education, p. 255-260; Special Session 2: Government Information, p. 261-282; Special Session 5 [sic.]: Geographic Information Systems, p. 281-294). Recordemos que aunque el DC sea un formato de propósito general, puede adecuarse para codificar cualquier tipo de información, gracias a su versatilidad y la posibilidad de establecer, además, perfiles de aplicación.

¹⁸⁵ La última versión (1.2.1.) de la especificación de metadatos IMS se publicó el 1 de octubre de 2001 y está basada en un esquema XML ([imsmd_rootv1p2p1.xsd](http://www.imsproject.org/metadata)). Se está trabajando también en la adaptación de este esquema a RDF. *Cfr.* <http://www.imsproject.org/metadata>

modelo DC, o formatos creados *ad hoc* como **GEM**¹⁸⁶ (*Gateway to Educational Materials*), **LOM** (*Learning Object Metadata*) surgido en el seno del *IEEE Learning Technology Standards Committee*, el esquema de metadatos europeo para la educación **EUN** (*European School Net*)¹⁸⁷ o el reciente esquema de metadatos, basado en RDF, **VTC** (*Virtual Teacher Centre*).

La información oficial es otro de los ámbitos que más ha cambiado a tenor de Internet. Los gobiernos de los distintos países están articulando servicios informativos para reactivar lo que ya se conoce como GOL (*Government On Line*) o *e-Government*, donde los metadatos están jugando un papel crucial en la organización y el acceso a la información gubernamental. En este sentido, destaca el formato **GILS** (*Government Information Locator Service*), uno de los modelos de metadatos más importantes en el contexto de la información pública americana. La base de datos del Servicio de Localización de Información Gubernamental (GILS), que responde a un proyecto de finales de 1994 para identificar los recursos de información pública en todo el gobierno federal, describe la información accesible de tales recursos y suministra ayuda en la recuperación de información. Los registros de metadatos del GILS describen los sistemas de información automatizada de las agencias federales americanas, de tal forma que se convierten en una herramienta fundamental para el acceso público a este tipo de información electrónica¹⁸⁸.

¹⁸⁶ Este modelo, como ya comentamos (*Vid.4.3.2.4.*), ha desarrollado su propio software de creación de metadatos GEMCat, así como su terminología controlada específica.

¹⁸⁷ Todos estos modelos, así como sus aplicaciones a proyectos concretos y las implicaciones de la información educativa en Internet, se han recopilado en una monografía en colaboración publicada en el año 2000. *Vid. Metadata and Organizing Educational Resources on the Internet*. Jane Greenberg, ed. New York, London, Oxford: The Harworth Information Press, 2000, 302 p.

¹⁸⁸ Para más información sobre este formato *Vid.* William E. Moen, Erin L. Stewart, Charles R. McClure. *The Role of Content Analysis in Evaluating Metadata for the U.S. Government Information Locator Service (GILS): Results from an Exploratory Study* [documento HTML]. En: *Paper index [of the] Second IEEE Metadata Conference, September, 16-17, 1997*. IEEE, 1997. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/wmoen/wmoen.htm> (consultado el 18 de junio de 1998).

Asimismo, distintos gobiernos nacionales han adaptado o desarrollado sus propios esquemas de metainformación. Destacamos, en este sentido, las iniciativas australiana (**AGLS**, *Australian Government Locator Service*) y neocelandesa (**NZGLS**, *New Zealand Government Locator Service*).

Finalmente, en el ámbito de la información geoespacial y de los Sistemas de Información Geográfica (GIS), sobresale el *Content Standard for Digital Geospatial Metadata* (CSDGM¹⁸⁹), más conocido como **FGDC**, que es la institución (*Federal Geographic Committee*) que creó en 1994, y mantiene y desarrolla desde entonces, este estándar de metadatos, que ha dado lugar a múltiples *clearinghouses*. Desde septiembre del 2000, esta norma *de facto* para los GIS se ha convertido en un borrador ISO (DIS) para su armonización con los estándares internacionales de codificación de la información geoespacial.

Como ya hemos mencionado, existen otros muchos modelos de metadatos aplicados a distintas áreas informativas¹⁹⁰, cuyo tratamiento no es el objetivo de esta investigación. No obstante, es preciso que destaquemos que en todos ellos, en mayor o menor medida, existe un esfuerzo de adaptación a la sintaxis XML. Los formatos específicos que en un principio (1995-1998) no eran más que un intento de reflejar la semántica descriptiva de una disciplina a través de un conjunto de etiquetas o elementos, sin embargo ahora están adoptando, cada vez más, la sintaxis flexible e integradora de XML. Quizás en los próximos años asistamos a una armonización de todos estos metadatos, pero hoy en día es preciso contemplar todas estas variedades que responden al principio de verticalización de la información y, por ende, al desarrollo de soluciones locales con vocación de acceso global. Es difícil prever qué modelos se convertirán en líderes dentro de su área, e incluso cuáles sobrevivirán independientes al desarrollo de la Web Semántica, de RDF y de la filosofía de

¹⁸⁹ Federal Geographic Data Committee. *Content Standard...* *Op. cit.*, <http://www.fgdc.gov/metadata/constan.html>

¹⁹⁰ *Cfr.* 3.4.

perfiles de aplicación del DC, pero hoy en día, todos estos formatos y otros muchos fundamentan proyectos de organización y recuperación de información electrónica temática.

5.2.3. RDF: un «metamodelo» de metadatos

El panorama de la metainformación que hemos reflejado hasta ahora, pone en evidencia que no hay un solo modelo de metadatos sino varios que responden a distintos escenarios informativos. Los formatos analizados no se excluyen entre sí, sino más bien, incorporan algunos elementos comunes (p. ej., title). Esta realidad ha hecho que se necesite también un marco para asegurar la coexistencia de esquemas de metadatos, primero a través del *Warwick Framework* (WF) y ahora a través de RDF (*Resource Description Framework*). RDF, más que un mero formato de metadatos es una infraestructura de descripción de recursos que trata de contener otros esquemas de metadatos en la misma línea difundida por el WF. De la misma forma que XML no es un lenguaje de marcado, sino un metalenguaje que permite definir otros lenguajes de estructuración de contenidos, RDF, desde nuestro punto de vista, no es un modelo de metadatos, sino un "metamodelo" que define una norma para crear modelos o esquemas de metadatos¹⁹¹.

Es el estándar más prometedor para asociar metainformación a recursos Web. Surgido, al igual que XML, en el seno del Consorcio World-Wide Web, RDF es una aplicación de metadatos que utiliza el Lenguaje de Marcado Extensible (XML) para describir un estándar para la interoperabilidad y el intercambio de DLOs. Es una manera de usar XML para datos (para metadatos) en lugar de sólo para documentos.

¹⁹¹ Para argumentar más sólidamente esta afirmación, podemos señalar que se trata de la forma de codificación de metainformación por excelencia en el nuevo contexto informativo protagonizado por XML y que conducirá a una Web más semántica. No en vano en el glosario especializado de Jensen, el término metadatos tiene asociada un ancla que remite automáticamente a la voz RDF. Bob Jensen. Resource Description Framework (RDF). En: *Jensen's Technology Glossary* [documento HTML]. San Antonio: Trinity University, rev. 27 de octubre de 2001. Disponible en: <http://www.trinity.edu/~rjensen/245glosf.htm#ResourceDescriptionFramework> (consultado el 3 de noviembre de 2001).

Como señala Miller¹⁹², *XML impone la necesidad de una restricción estructural para proporcionar métodos inequívocos de expresión semántica*. RDF no es más que la infraestructura que permite esa restricción gracias a la codificación, intercambio y reutilización de metadatos estructurados. Con estas prerrogativas, interoperabilidad y estructuración, que como venimos demostrando en esta investigación parecen ser las características fundamentales de un formato de metadatos coherente con la realidad de la información Web, no es arriesgado decir que RDF asegura una búsqueda y recuperación mejores en la Web Semántica proyectada por Berners Lee.

RDF surge, para Hjelm¹⁹³ y también para otros autores, de la intersección del mundo de la gestión y representación del conocimiento y del mundo de los metadatos en bibliotecas. Se nutre de otros trabajos previos, como PICS (*Platform for Internet Content Selection*) y el Dublin Core/Warwik Framework (que, como vimos vincula su desarrollo actual a RDF), así como de la actividad en materia de metadatos de los principales vendedores de software para Internet (MCF, CDF). El objetivo del formato RDF es proporcionar una arquitectura flexible para gestionar conjuntos específicos de aplicaciones de metadatos que pueden procesarse por ordenador. Además, RDF permitirá la integración de firmas digitales, lo que posibilitará que las organizaciones obtengan la confianza de los usuarios al firmar los paquetes de metadatos como descripciones auténticas y "honestas" de los recursos de la Red. Como se puede observar, la finalidad del *Resource Description Framework* va más allá de la aplicabilidad para una mejor estructuración semántica de la información y para una recuperación más precisa, que es la aplicación de los metadatos que tratamos de demostrar en esta investigación. Este formato permite también otras utilidades de los metadatos, como la posibilidad de valoración de la calidad del

¹⁹²[...] *XML that imposes needed structural constraints to provide unambiguous methods of expressing semantics*. Eric Miller. An introduction to the Resource Description Framework [documento HTML]. *D-Lib Magazine*, May 1998. Disponible en: <http://www.dlib.org/dlib/may98/miller/05miller.html> (consultado el 18 de julio de 1998).

¹⁹³ Johan Hjelm. *Op. cit.*, p. 2.

contenido que es otra de las necesidades que tiene el usuario al enfrentarse a la información heterogénea de la Web¹⁹⁴. Otro aspecto destacable es su aptitud para promover un medio de edición de vocabularios que sean, al mismo tiempo, legibles por el hombre y procesables por el ordenador, como forma de exhortar la reutilización y la extensión de la semántica de los metadatos entre las diferentes comunidades de información de Internet. Por ejemplo, vocabularios como RSS¹⁹⁵ (*RDF Site Summary*), algo así como la evolución lógica del CDF de Netscape a tenor de RDF, que se ha convertido en el formato de sindicación de contenidos¹⁹⁶; o PRISM (*Publishing Requirements for Industry Standard Metadata*), que define un vocabulario XML para el intercambio y preservación de contenidos y metadatos de revistas, catálogos, libros, noticias, etc. para las organizaciones que producen y diseminan información. RDF está en proceso de desarrollo e implementación en diversos colectivos que utilizan e intercambian información electrónica, pero de todos los usos posibles de RDF¹⁹⁷, nos interesa su aptitud para el intercambio de información en bibliotecas digitales.

Como afirma el propio inventor de la WWW¹⁹⁸, *la introducción del RDF no ha sido clara ni fácil, y se ha discutido mucho acerca de cómo debería ser introducido y*

¹⁹⁴ Al hablar de las aplicaciones de los metadatos —apartado 3.4, de esta tesis—, comentamos que Dempsey y Heery señalaron que una de las necesidades del usuario adulto era el conocimiento del contenido de los recursos a los fines de determinar su adecuación para los menores. En este sentido el formato RDF, como heredero, en cierta medida, del modelo de *Plataforma de Selección de los Contenidos de Internet* (PICS) desarrollado también por el W3C en 1995, permite valorar los contenidos "honestos" o adecuados gracias a la integración de las firmas digitales.

¹⁹⁵ Vid. Capítulo 4, nota 152.

¹⁹⁶ Vid. Glosario, *syndication/syndication format*.

¹⁹⁷ La complejidad y la proyección de RDF en distintos colectivos que intercambian información daría lugar a tesis o trabajos independientes con proyectos de aplicabilidad individuales. Aquí sólo tratamos de presentar RDF como un metamodelo de metadatos para las bibliotecas digitales.

¹⁹⁸ Tim Berners Lee. *Tejiendo... Op. cit.*, p. 167.

hasta si debería serlo¹⁹⁹. Sin embargo, el modelo y la sintaxis de RDF²⁰⁰ son, desde febrero de 1999, una Recomendación del Consorcio Web y el esquema RDF²⁰¹ es Recomendación Candidata desde marzo de 2000. Resulta difícil sintetizar aquí toda la complejidad, importancia, funcionalidad y potencialidad del *Resource Description Framework*. No obstante, partiendo de la documentación básica relativa a RDF, convertida en especificación de acceso público por el W3C, se pueden destacar tres aspectos de la semántica funcional de este metamodelo de metadatos: un modelo de datos, una sintaxis y un esquema, que resumimos a continuación y que seguiremos reflejando en otras partes de este trabajo.

a) El modelo de datos y sintaxis RDF

En esencia, RDF es un modelo para la representación de las denominadas propiedades y los valores de esas propiedades. El modelo RDF se constituye sobre principios bien establecidos en el DC. Las propiedades de RDF se pueden entender como atributos de los recursos y, en este sentido, corresponden a los pares tradicionales de atributo-valor que hemos visto en otros formatos. Además estas propiedades, representan las relaciones entre los distintos recursos de información, de

¹⁹⁹ Hemos asistido a muchos de estos debates en torno a la incorporación de RDF a través de la suscripción a la lista RDF-interest, desde su creación en agosto de 1999. Mensajes públicos de la lista en: <http://lists.w3.org/Archives/Public/www-rdf-interest>

²⁰⁰ RDFMS, *Op. cit.*, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222> (traducción al español de Eva Méndez en: <http://www.sidar.org/traduc/rdfesp.htm>). Para una interpretación exhaustiva de esta recomendación, *Vid.* Kal Ahmed, et al. *Op. cit.*, 97-164.

²⁰¹ [RDFS] World Wide Web Consortium. *Resource Description Framework (RDF) Schema Specification 1.0 W3C Candidate Recommendation 27 March 2000* [documento HTML]. Dan Brickley R.V. Guha, eds. W3C, 27 de marzo de 2000. Disponible en: <http://www.w3.org/TR/2000/CR-rdf-schema-20000327> (consultado el 8 de abril de 2000). (También hemos realizado la versión española de esta recomendación candidata: <http://www.sidar.org/traduc/rdfsch.htm>). Para una interpretación exhaustiva sobre el *schema* RDF, *Vid.* Kal Ahmed, et al. *Op. cit.*, 167-201.

modo que este modelo puede parecer un esquema entidad-relación²⁰² de las bases de datos relacionales tan conocidas en el mundo bibliotecario en los SIGB. El modelo RDF se nutre también del diseño orientado a objetos, donde los recursos corresponden a objetos y las propiedades corresponden a ejemplos de variables. No en vano venimos resaltando desde el principio que el método de asignación de metadatos para describir recursos de información se basa en una unidad documental que entiende al documento como un objeto (DLO).

Según lo expresado, el modelo de datos que propone RDF consiste en tres tipos de objetos o triples (Fig. 53):

- *Recursos*: Cualquier objeto Web identificable unívocamente por un URI (es decir, un identificador uniforme de recursos como un URL). Un recurso puede ser un documento HTML; una parte de una página Web, como por ejemplo un elemento HTML o XML dentro de un documento fuente; una colección de páginas; un sitio Web completo; y, en síntesis, cualquier recurso entendido como objeto de información.
- *Tipos de propiedades (Property Types)*: Las propiedades son atributos o relaciones asociadas a los recursos. Cada tipo de propiedad tiene sus valores correspondientes y define su significado específico, los tipos de objetos que se pueden describir con una propiedad determinada, así como las relaciones que existen entre las distintas propiedades.
- *Descripciones o sentencias (Statements)*: Son el conjunto de un recurso, un nombre de propiedad y el valor de esa propiedad. Estas tres partes individuales de una sentencia se denominan, respectivamente, sujeto, predicado y objeto. El

²⁰² Para una comparación del modelo genérico que subyace a todos los formatos de metadatos con los modelos entidad-relación, *Vid.* Tim Berners-Lee. Metadata Architecture: Document, Metadata and Links [documento HTML]. *Design Issues*. World-Wide Web Consortium, 6 de enero de 1997, rev. 6 de febrero de 1998. Disponible en: <http://www.w3.org/DesignIssues/Metadata.html> (consultado el 6 de julio de 1998).

objeto de una sentencia (el valor de la propiedad) puede ser otro recurso o un literal, es decir, un recurso (especificado por un URI), o una cadena simple de caracteres u otros tipos de datos definidos por XML.

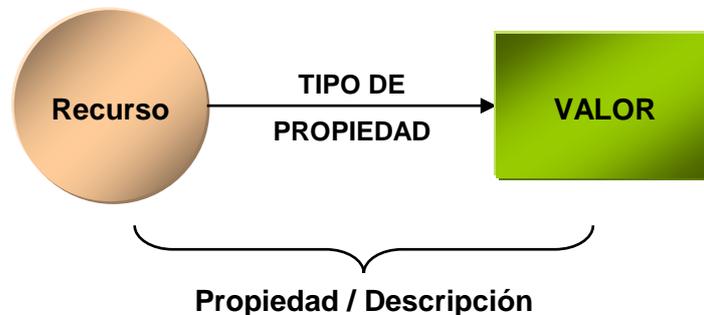


Fig. 53. Expresión gráfica del modelo de datos RDF

Un ejemplo de este modelo, continuando con el supuesto de esta tesis como si fuese un documento Web, sería:

De la sentencia: "Eva Méndez es el autor del recurso: <http://www.bib.uc3m.es/~mendez/tesis/tesis.htm>", se desprenderían los siguientes objetos:
 Recurso: <http://www.bib.uc3m.es/~mendez/tesis/tesis.htm>
 Tipo de propiedad: author, creator
 Valor: Eva Méndez

Por otra parte, la sintaxis que representa este modelo tendrá que almacenar las sentencias o descripciones que se desprenden de él, de forma legible por ordenador. Para ello, la sintaxis elegida para el modelo RDF es el lenguaje XML, por su flexibilidad y propósito más general que el de HTML, de tal forma que XML soporta una representación consistente de la semántica que se genera a partir del conjunto de reglas que facilita RDF.

Con frecuencia, el mayor problema es la disparidad terminológica ya que, en diferentes comunidades de usuarios (o en diferentes proyectos de biblioteca digital), los tipos de objetos pueden tener diferente denominación. Por ejemplo, a la hora de llevar a una sintaxis de metainformación la sentencia arriba expuesta —"Eva Méndez es el autor del recurso <http://www.bib.uc3m.es/~mendez/tesis/tesis.htm>"—, el modelo de metadatos de la TEIH, vendría mediatizado por el tipo de propiedad Author; por el contrario, la misma propiedad en el modelo del DC se definiría por el tipo de

propiedad `Creator`. Esta diversidad se soluciona en RDF utilizando el elemento *namespace*²⁰³ o espacio de nombre (`xmlns`) del lenguaje XML, que permite expresar un espacio o esquema inequívoco al consignar el recurso que define la semántica correspondiente al comienzo de una declaración o registro de metadatos. Por ejemplo, si los esquemas que se siguen para el registro de metadatos relativo al documento Web de Chris Taylor *Introduction to Metadata* que hemos citado, fuesen el DC y el AGLS, se expresarían de la siguiente forma:

```
<? xml version="1.0" ?>
<RDF xmlns="http://w3.org/TR/1999/PR-rdf-syntax-19990105#"
      xmlns:DC="http://purl.org/dc/elements/1.1#"
      xmlns:AGLS="http://naa.gov.au/AGLS#">
  <Description about
    ="http://www.library.uq.edu.au/iad/ctmeta4.html">
    <DC:Title> An introduction to metadata </DC:Title>
    <DC:Creator> Crys Taylor </DC:Creator>
    <DC:Date> 1999-04-01 </DC:Date>
    <DC:Subject> Metadata, RDF, Dublin Core </DC:Subject>
    <AGLS:Function> Information Management - Internet
  </AGLS:Function>
  </Description>
</RDF>
```

Según esto, quedan definidos ambos esquemas: *Resource Description Framework* como mecanismo para expresar el modelo de metadatos RDF, y los esquemas del *Dublin Core* y del *Australian Government Locator Service* como vocabularios de designación de tipos de atributos; de modo que estas declaraciones del *namespace* más sus respectivos URI (URL) definen los esquemas correspondientes. Mediante el empleo de los espacios de nombre, RDF permite utilizar varios conjuntos de elementos de metadatos de forma modular, eliminando las colisiones entre términos utilizados por distintas comunidades, y permite asociar un prefijo a cada vocabulario, eliminando, de este modo, la ambigüedad en el caso de que varios vocabularios tuviesen algún término en común.

En el caso del DC, RDF proporciona una sintaxis de almacenamiento muy parecida, —con las correspondientes salvedades debido a la naturaleza de las

²⁰³ Vid. Glosario.

relaciones y a que se trata de un lenguaje legible por el hombre además de por máquina— a lo que el MARC hace con las Reglas de Catalogación. A pesar del desarrollo convergente de ambos estándares, RDF no se limita a representar DC, y el Dublin Core no sólo se representa en RDF, como hemos visto²⁰⁴. A diferencia del formato MARC, RDF tiene también mecanismos para referenciar una relación de una norma en otras. La finalidad es, pues, que una aplicación no esté subyugada a la comprensión de un estándar particular o una versión del mismo.

Como en los otros formatos que hemos descrito, la comprensión del modelo mejorará si aportamos un ejemplo real. En este caso, comparamos la codificación de metadatos de uno de los borradores del RDFMS-1998²⁰⁵ en el proceso de desarrollo de la recomendación, con el código fuente de la recomendación del modelo y la sintaxis que venimos refiriendo (RDFMS-1999)²⁰⁶:

Ejemplo RDF (1)

a) Registro de metadatos del *Working Draft* del RDFMS de 19 agosto de 1998²⁰⁷:

²⁰⁴ A pesar de que esta afirmación es cierta en sentido estricto, el desarrollo de DCMI a la par que el de RDF hace que hoy por hoy no se entienda la semántica del DC sin la representación RDF, y viceversa. Así, por ejemplo, lo refleja el capítulo 12 del manual de XML de St. Laurent y Biggar, titulado: *Organizing Information: RDF and Dublin Core*. Simon St. Laurent, Robert Biggar. *Op. cit.*, p. 207-228.

²⁰⁵ World Wide Web Consortium. *Resource Description Framework (RDF): Working Draft 19 August 1998* [documento HTML] Ora Lassila and Ralph R. Swick, eds. W3C, 19 de agosto de 1998. Disponible en: <http://www.w3.org/TR/1998/WD-rdf-syntax-19980819> (consultado el 24 de agosto de 1998).

²⁰⁶ RDFMS. *Op. cit.*, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>

²⁰⁷ En el documento de trabajo que estamos refiriendo a través de su declaración de metadatos, se dice expresamente que: *el código HTML fuente de este documento contiene RDF embebido por ello no será validado por la DTD actual [1998] del HTML 4.0.* —la versión vigente de HTML, en agosto de 1998 era: <http://www.w3.org/TR/1998/REC-html40-19980424>— *Una solución para esta necesidad de validación del estilo-DTD puede estar mediatizada por el trabajo del W3C en un futuro. Empero no estar aceptado todavía por la última versión del HTML, no es óbice para poder incluirlo en la cabecera <HEAD> . . . </HEAD>.*

```

<HEAD>
<TITLE>Resource Description Framework (RDF) Model and Syntax
Specification</TITLE>
<META NAME="Creator" CONTENT="Ora Lassila, Ralph Swick"/>
<rdf:Description about=""
xmlns:rdf="http://www.w3.org/TR/WD-rdf-syntax#"
xmlns:dc="http://purl.org/metadata/dublin_core#"
xmlns:ddc="http://purl.org/net/ddc#"
dc:Title="Resource Description Framework (RDF) Model and
Syntax Specification"
dc:Description="The Resource Description Framework (RDF) is
a foundation for
processing metadata; it provides interoperability between
applications that exchange machine-understandable information
on the Web. RDF emphasizes facilities to enable automated
processing of Web resources."
dc:Publisher="World Wide Web Consortium"
dc>Date="1998-08-19"
dc:Format="text/html"
dc:Type="technical specification"
dc:Language="eng">
<dc:Subject resource="http://purl.org/net/ddc/025.30285"
ddc:Class="025.30285"
ddc:Heading="data processing computer applications"/>
<dc:Subject resource="http://purl.org/net/ddc/025.316"
ddc:Class="025.316"
ddc:Heading="Machine-readable catalog record formats"/>
<dc:Subject
ddc:Class="025.302855741"
ddc:Heading="Applications of computer file organization
and access methods"/>
<dc:Creator>
<rdf:Bag rdf:_1="Ora Lassila" rdf:_2="Ralph Swick"/>
</dc:Creator>
</rdf:Description>
</HEAD>

```

El hecho de no ser un ejemplo ficticio sino extraído de un código fuente real, explica que tenga los elementos que estaban aceptados en la especificación del HTML 4.0 <http://www.w3.org/TR/1998/REC-html40-19980424> como las etiquetas <TITLE> y <META> en la cabecera; y también que no contenga etiquetas <RDF:RDF> . . . </RDF:RDF> que indiquen el principio y el final de la descripción dentro de un documento XML. Además, podemos señalar que:

- En este ejemplo particular, se observan tres espacios de nombre (`xmlns`) definidos por su correspondiente dirección que identifica el vocabulario en el que se basan: por un lado *Resource Description Framework* (`rdf`), por otro, el Dublin Core (`dc`) y, finalmente, el `ddc` (que corresponde a la *Decimal Dewey Classification* como esquema de clasificación).

- El elemento `<rdf:description>` basado en el esquema RDF (RDFS), se utiliza para identificar el recurso que se está describiendo que, en este caso, es además, implícitamente, la declaración del esquema.

El resto de los elementos son los del modelo DC y del DDC, salvo el elemento `rdf:bag_n` que se usa para declarar la propiedad intelectual del documento en el caso que haya varios autores, donde n es el un valor individual para los autores, cuyo orden no tiene importancia porque comparten la propiedad intelectual del recurso.

- b) Registro de metadatos del documento de la Recomendación del RDFMS del 22 de febrero de 1999:

```
<HEAD>
<TITLE>Resource Description Framework (RDF) Model and Syntax
Specification</TITLE>
<LINK rel="meta" href="http://www.w3.org/TR/1999/REC-rdf-
syntax-19990222/metadata">
<STYLE TYPE="text/css">
.EXAMPLE { margin-left: 1em }
</STYLE>
</HEAD>
```

A diferencia del caso anterior, donde se utilizaba una sintaxis abreviada de RDF embebida en el documento HTML 4.0, ahora se utiliza el elemento LINK de la cabecera en HTML 4.01 para relacionar el fichero de metadatos almacenado aparte `<http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/metadata>` donde la codificación RDF es la siguiente:

```
<?xml version="1.0" ?>
<rdf:Description about="Overview.html"
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/metadata/dublin_core#" (208)
xmlns:ddc="http://purl.org/net/ddc#"
dc:Title="Resource Description Framework (RDF) Model and Syntax
Specification" dc:Description="The Resource Description
```

²⁰⁸ Este es el URL tal y como aparece en el documento de metadatos que estamos citando. Sin embargo, el URI para definir el espacio del nombre DCMI es: <http://purl.org/dc/elements/1.1>, que a su vez reenvía al URI <http://dublincore.org/2001/08/14/dces#> donde está definida formalmente la declaración del *Schema* RDF para el conjunto de elementos DC, con fecha 14 de agosto de 2001. *Vid. Namespace Policy for the Dublin Core Metadata Initiative* (DCMI Recommendation, 15 de noviembre de 2001) `<http://dublincore.org/documents/2001/10/26/dcmi-namespace>`.

```

Framework (RDF) is a foundation for processing metadata; it
provides interoperability between applications that exchange
machine-understandable information on the Web. RDF emphasizes
facilities to enable automated processing of Web resources."
dc:Publisher="World Wide Web Consortium" dc>Date="1999-02-22"
dc:Format="text/html" dc:Type="technical specification"
dc:Language="en">
  <dc:Subject rdf:resource="http://purl.org/net/ddc/025.30285"
ddc:Class="025.30285" ddc:Heading="data processing computer
applications" />
  <dc:Subject rdf:resource="http://purl.org/net/ddc/025.316"
ddc:Class="025.316" ddc:Heading="Machine-readable catalog
record formats" />
  <dc:Subject ddc:Class="025.30285741"
ddc:Heading="Applications of computer file organization and
access methods" />
  <dc:Relation resource="http://www.w3.org/TR/REC-rdf-syntax" />
<dc:Creator>
<rdf:Bag rdf:_1="Ora Lassila" rdf:_2="Ralph Swick" />
</dc:Creator>
</rdf:Description>

```

A continuación, y de forma paralela a lo que hicimos al hablar del DC y de su representación en HTML, utilizamos el supuesto de que esta tesis es un documento para la Web al que asignamos un registro de metadatos DC codificado, en este caso, en RDF:

Ejemplo RDF (2): DC en sintaxis RDF

```

<?xml version = "1.0"?>
<RDF xmlns="http://www.w3.org/TR/WD-rdf-syntax#">
<DC xmlns="http://metadata.net/dstc/DC-10-EN">
<RDF:RDF>
<RDF:Description
xml:lang="es"about="http://www.bib.uc3m.es/~mendez/tesis/tesis.htm">
<DC:Title> Metadatos y recuperación de información: estándares,
problemas y aplicabilidad en bibliotecas digitales</DC:Title>
<DC:Creator.PersonalName>Eva M. Méndez Rodríguez
</DC:Creator.PersonalName>
<DC:Creator.PersonalName.Address>emendez@bib.uc3m.es</DC:Creator.Per
sonalName.Address>
<DC:Subject> metadatos, recuperación de información, bibliotecas
digitales, estándares, normas, Biblioteca Virtual Cervantes,
sistemas de información de calidad, estructuración de la
información, metainformación </DC:Subject>
  <DC:Description>Tesis doctoral donde se trata de mostrar la
utilidad de los metadatos para la recuperación de información, sobre
todo en contextos finitos de información o bibliotecas digitales que
permiten diseñar sistemas de información de calidad. Se estudian
distintos formatos, normas y problemas relacionados con este tema.
además se plantea un proyecto de aplicación de metadatos a la
Biblioteca Virtual Miguel de Cervantes</DC:Description>
<DC:Publisher.CorporateName.Address>Universidad Carlos III de Madrid
</DC:Publisher.CorporateName.Address>
<DC:Contributor.PersonalName> José A. Moreiro González
</DC:Contributor.PersonalName>

```

```

<DC:Date.Accepted DC:Scheme="ISO8601"> 2001-12-13
</DC:Date.Accepted>
<DC:Type> Tesis </DC:Type>
<DC:Format DC:Scheme="IMT"> text/html</DC:Format>
<DC:Identifier DC:Scheme="URI">
http://www.bib.uc3m.es/~mendez/tesis/tesis.htm </DC:Identifier>
<DC:Language> es </DC:Language>
<DC:Relation.IsReferencedBy>
http://www.bib.uc3m.es/~mendez/profesional/index.htm
</DC:Relation.IsReferencedBy>
<DC:Rights> Eva Méndez </DC:Rights>
</RDF:Description></RDF:RDF>

```

Una vez creado este registro de metadatos, tendríamos que vincularlo al documento Web en cuestión, pudiendo darse dos circunstancias:

- Que nuestro documento Web fuese un documento XML, en cuyo caso se podría embeber directamente esta cabecera en el código fuente.
- Que nuestro documento Web estuviese en HTML 4.01: en este supuesto, se podría incluir en la cabecera del documento HTML con una sintaxis abreviada — como en el Ejemplo RDF (1), a)—, o también de forma externa al documento, utilizando la etiqueta <LINK> del HTML —como en el ejemplo RDF (1) b)—, p. ej., así:

```

<LINK rel="meta"
href="http://www.bib.uc3m.es/~mendez/tesis/tesis.rdf">

```

b) El esquema RDF

El otro elemento clave de la semántica funcional de RDF es el esquema o esquemas (RDFS). *Los esquemas²⁰⁹ RDF son algo parecido a las DTDs en XML, pero tienen su propia sintaxis y sus propias particularidades²¹⁰.* Una DTD de XML proporciona las restricciones específicas en la estructura de un documento, mientras que un esquema RDF proporciona información sobre la interpretación dada en un modelo de datos RDF. A pesar de la flexibilidad de la sintaxis XML en RDF, un solo

²⁰⁹ Cfr. Glosario *schema*, *scheme*.

²¹⁰ Simon St. Laurent, Robert Biggar. *Op. cit.*, p. 221.

esquema sintáctico no es suficiente para los propósitos de este último. Se han de especificar también las restricciones que deben seguir estos modelos de datos, las cuales se establecen a través del esquema. Además, los esquemas RDF son diferentes a las DTD de XML porque, en vez de modelos de contenido y atributos, RDFS define clases de jerarquías y restricciones de uso para esas clases. También se diferencia del esquema XML fundamentalmente porque éste proporciona un modelo para describir la estructura de la información en toda una clase de documentos, reemplazando así, como apuntamos anteriormente, a las DTDs. No obstante, un RDFS proporciona un mecanismo para declarar las propiedades de los elementos y las relaciones entre las propiedades y los recursos, aportando además una semántica para todo ello, esto es una forma de crear vocabularios.

El esquema RDF se utiliza para declarar vocabularios (como por ejemplo RSS), esto es, conjuntos de tipos de propiedades definidos por una comunidad. Así un esquema RDF se puede considerar metadatos para crear otro metadatos (sean estos basados en el recurso, o basados en materias²¹¹). Si la sintaxis y el modelo RDF pueden considerarse de propósito general, la facilidad de crear vocabularios a través del RDFS nos permite considerarlo un estándar de metainformación de propósito específico, o mejor, tratando de evitar las denominaciones que hemos utilizado para clasificar las distintas iniciativas de metainformación, es un metamodelo de metadatos. Por eso podemos decir que es más que XML²¹² y más que un mero

²¹¹ Según la tipología de metadatos señalada por Kashyap y Sheth y Kal Ahmed, et al. *Vid.* 3.5.

²¹² Casi todos los autores que tratan RDF hacen alguna comparación entre éste y XML, *V.gr.* Ora Lassila, Norm Medeiros, Hjelm, etc. (*Cfr.* Bibliografía). Nos parece especialmente elocuente la afirmación de Dumbill cuando dice que: *RDF es a XML lo que Perl es a C++; RDF es de hecho mucho más 'Webby' que el propio XML* A pesar de que en el contexto de la frase de Dumbill se puede atisbar el significado de *Webby*, (*Cfr.* Glosario) se refiere a que RDF es más adecuado para la Web que XML en lo que a codificación de metadatos se refiere. Edd Dumbill. *RDF - Why we Should Care - and RSS* [documento HTML]. *Edd Dumbill Weblog. O'Reilly*, 6 de marzo de 2000. Disponible en: [http://weblogs.oreillynet.com/edd/discuss/msgReader\\$65](http://weblogs.oreillynet.com/edd/discuss/msgReader$65) (consultado el 8 de julio de 2001). Afirmaciones como ésta, nos hacen pensar que todos los modelos de metadatos de propósito específico (5.2.2), creados originalmente en forma de DTDs de SGML pero que se han reformulado en la mayoría de los casos como DTDs de XML, se reformularán definitivamente como esquemas RDF.

modelo de metadatos. No obstante el esquema RDF es todavía una recomendación candidata del W3C²¹³ y quedan muchos aspectos por perfilar y clarificar.

Entre los múltiples debates que ha generado este formato de metadatos en la comunidad de la información, es importante el relacionado con los protocolos de búsqueda. En una discusión acerca de las ventajas de RDF frente a otros estándares similares de intercambio de información, desde los inicios de este formato (1998) se ha reconocido (Andrew Powell, Dan Brickley, Brian Kelly, Renato Iannella) *que la idea básica que subyace al RDF es similar a la que hay detrás del formato GRS.1 (Generalized Record Syntax) en Z.39.50, es decir, un árbol jerárquico de información dividida en identificadores etiquetados, donde el significado y el formato (marcado descriptivo y de procedimiento) de las etiquetas viene definido por un conjunto de etiquetas importadas (GRS.1) o esquemas (RDF)*²¹⁴.

El solapamiento con Z39.50 es ligeramente mayor cuando la comunidad Web ve RDF como el fundamento para la búsqueda a través de distintos dominios en Internet. Con esto podemos decir que RDF puede concebirse también como una superación del sistema de búsqueda de información Z39.50. En el inicio del desarrollo de RDF, se resaltó también (Dan Brickley) que podría ser *lamentable* convertir ahora los datos basados en Z39.50 o formato MARC en RDF, en el sentido de que compañías poderosas, como Microsoft e IBM, probablemente vean RDF como el futuro de esta clase de aplicaciones de búsqueda en entornos heterogéneos de información, más que las aplicaciones basadas en Z39.50. Por ello, en el seno de Mozilla se inició en 1999 un proyecto de integración entre RDF y Z39.50²¹⁵.

²¹³ RDFS, *Op. cit.*, <http://www.w3.org/TR/2000/CR-rdf-schema-20000327>

²¹⁴ Matthew J. Dovey. UKOLN RDF Seminar, Bath [documento HTML]. *Ariadne*, issue 15-May-1998, rev. 17 de mayo de 1998. Disponible en: <http://www.ariadne.ac.uk/issue15/events/stakis.html> (consultado el 21 de julio de 1998).

²¹⁵ Dan Brickley (daniel.brickey@bristol.ac.uk). ANNOUNCE: Mozilla RDF/Z39.50 Integration Project [mensaje en newsgroup]. *Netscape.public.mozilla.rdf*, 24 de agosto de 1999 01:04:50 GMT (Consultado el 7 de septiembre de 1999) y Dan Brickley. Mozilla RDF / Z39.50 Integration Project [cont.]

Así, en el año 2000, en el encuentro del ZIG celebrado en Washington, se acordó revisar el estándar, aprovechando los cinco años que habían pasado desde su creación, con propuestas que iban desde su reformulación en XML (ZML) a tendencias aún más drásticas, como la de John Kunze que pensaban en comenzar de nuevo. Es destacable la dura e irónica defensa de la perdurabilidad de Z39.50 realizada por Taylor en dicho encuentro del ZIG. Este autor afirma expresamente que *Z39.50 tal y como se conoce (2000) proporciona una funcionalidad que la "Mafia Web" no ha conseguido duplicar todavía*²¹⁶. Por "mafia Web", entiende todas las actividades del W3C en materia de normalización de estructuración de la información (XML, RDF, SOAP, XQL, etc...). A pesar de la dureza de Taylor sobre las iniciativas del W3C, es innegable hoy en día, como demostraremos en el capítulo siguiente, la capacidad de este consorcio para generar estándares *"de facto formales"*²¹⁷. Por todo ello, la evolución lógica de este protocolo internacional (ISO 23950), es compatibilizar la sintaxis de registros XML con la de GRS.1, lo que ha conducido, a raíz del encuentro del ZIG de octubre de 2001, y sin que ello obste para otras propuestas, a un nuevo borrador de la norma americana²¹⁸.

[documento HTML]. The Mozilla Organization, 22 de agosto de 1999. Disponible en: <http://www.mozilla.org/rdf/doc/z3950.html> (consultado el 7 de septiembre de 1999).

²¹⁶ Mike Taylor. *Z39.50: The Empire Strikes Back* [documento HTML]. Mike Taylor, 6 December 2000. Disponible en: <http://www.miketaylor.org.uk/tech/esb.html> (consultado el 8 de diciembre de 2001).

²¹⁷ Esta paradójica denominación de las normas Web, la aclararemos en el capítulo siguiente, al hablar de las *Public Access Specifications* (PAS) con carácter eminentemente normativo, que emana el W3C. Cfr. 6.2.

²¹⁸ *ANSI/NISO Z39.50-2001, Op. cit.*, <http://www.loc.gov/z3950/agency/revision/part1.pdf> y <http://www.loc.gov/z3950/agency/revision/part2.pdf>

Otra de las cuestiones básicas sobre RDF es el nivel de implantación que tiene este modelo en el mercado del software, esto es, qué aplicaciones y herramientas soportan este formato, tanto para la creación y validación de metadatos conforme a su modelo y sintaxis, como para la generación de nuevos esquemas RDF o para el

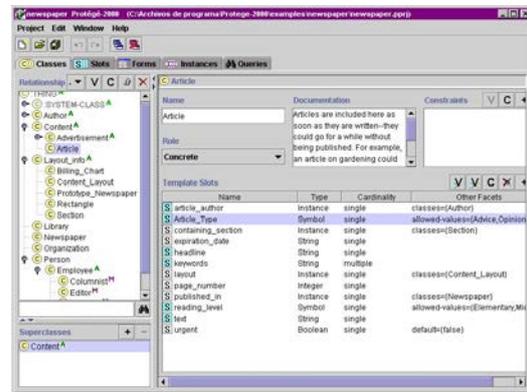


Fig. 54. Protégé 2000

almacenamiento y búsqueda de estos datos. En cuanto a la edición de metadatos, ya explicamos en el capítulo anterior algunas herramientas que permitían crear metadatos en RDF (DC-Dot Reggie, y Metabrowser; este último permitía no sólo editar metadatos sino también enviarlos a una base de datos). Sin embargo, RDF requiere herramientas más complejas (APIs, *parsers*, etc.)²¹⁹ como por ejemplo, Protégé 2000²²⁰ (Fig. 54.), que utiliza RDF como formato de almacenamiento pero permite también generar esquemas y vocabularios RDF.

Sin perjuicio de lo atinado de la clasificación de los modelos de metadatos en bandas o zonas de Dempsey y Heery —a la que venimos aludiendo a lo largo esta investigación a pesar de su carácter "obsoleto" (1997)— RDF no ha sido contemplado en ella por tratarse obviamente de un estándar desarrollado posteriormente (1998-99). Aún así, basándonos en los criterios de clasificación

²¹⁹ Para un completo listado de todas estas herramientas clasificadas. *Vid.* Johan Hjelm. *Op. cit.*, p. 255-267.

²²⁰ Programa-versión: Protégé 2000

Homepage: <http://protege.stanford.edu/index.shtml>

Tipo: Freeware

Plataforma: Windows, MacOS X, Aix, Solaris, Linux, etc.

Nombre (tamaño) del fichero: install_protege.exe (12.551 Kb)

Descarga:

http://www.smi.stanford.edu/projects/protege/download/release/InstData/Windows/VM/install_protege.exe

Fecha de descarga: 29/03/2001 15:38

empleados por estos autores, podemos decir que este modelo debería constituir una zona independiente, que vamos a denominar "banda 0", por ser una infraestructura contenedora del resto de formatos o un metamodelo de metadatos, como lo hemos definido aquí, que persigue la interoperabilidad entre todos ellos. Su desarrollo está mediatizado por la madurez de la arquitectura Web como sistema multidisciplinar de información. No obstante, comparte algunas características con las zonas de metadatos establecidas por Dempsey y Heery:

- Al estar vinculado al desarrollo del Dublin Core, teniendo en cuenta además su sencillez de designación, el hecho de no estar relacionado con ninguna área específica de conocimiento y de estar especialmente abocado a la recuperación de recursos en Internet, son características que acercan este formato a la banda 2.
- Por otra parte, el estar fundamentado en XML, la posibilidad de expresar gran riqueza funcional y gran variedad de relaciones semánticas, así como su adecuación para presentar estructuras jerárquicas complejas o describir colecciones de documentos, son características que nos podrían hacer pensar en su pertenencia a la banda documental.

Aunque RDF completa el amplio espectro de soluciones disponibles para desarrollar metadatos en la Web que hemos querido reflejar en este capítulo, sus objetivos son más amplios y las oportunidades potenciales que ofrece son enormes. Esto nos conduce a augurar su éxito e implantación²²¹.

Las aplicaciones o proyectos actuales del RDF no sólo responden a contextos de investigación o universitarios, como veíamos en los otros dos formatos de metadatos que se orientaban a intereses individuales de colectivos concretos de información, sino que su uso e implantación están estrechamente vinculados al

²²¹ En este caso, existe además un entusiasmo implícito por parte de fuertes empresas, como Netscape o Microsoft, en el desarrollo y adopción de esta infraestructura de descripción de recursos: Netscape 6.X (Mozilla) usa RDF para gestionar mapas de los sitios Web y listados de *bookmarks* e históricos. Más información en: <http://www.mozilla.org/rdf/doc>

desarrollo del nuevo y complejo concepto de Web Semántica, donde RDF es una estructura para construir lenguajes lógicos que puedan funcionar juntos en esa nueva Web. Como expresa Tim Berners Lee²²²:

Si HTML y la Web hicieron que todos los documentos en línea parecieran un solo libro enorme, el RDF y los lenguajes de esquema o deducción harían que todos los datos del mundo parecieran una enorme base de datos.

En esta afirmación del inventor de la Web se encierra toda la complejidad que advertimos en RDF, así como toda la potencialidad que este formato tiene para la representación y organización del conocimiento. El hecho de que pueda formar una "enorme base de datos" en la futura Web Semántica, nos da la pista para, de momento, aplicarlo a las bases de datos finitas y/o locales de distintas bibliotecas digitales²²³, sólo así en un futuro se podrá aspirar a la integración de todas ellas.

Para concluir este capítulo en torno a los modelos y/o esquemas de metainformación debemos apuntar, interpretando a Marcoux y Sévigny²²⁴ que, a pesar de la plétora de formatos existente en el mundo de la información, en muchos casos efímeros e incompatibles, la información Web no satisface las necesidades de acceso, diseminación y conservación que caracterizan la información documental. Por ello, debemos prestar atención a la necesidad de normalización de estos formatos de tal forma que se puedan diseñar sistemas de transferencia de información basados en ellos. El uso eficaz de metadatos requiere convenciones o normalizaciones relativas a la semántica, sintaxis y estructura de la metainformación. Por esta razón, dedicaremos el capítulo siguiente al aspecto de la normalización, y en el capítulo 8 hablaremos de la interoperabilidad entre modelos, sistemas y estándares de metadatos.

²²² Tim Berners Lee. *Tejiendo... Op. cit.*, p. 171.

²²³ *Cfr.* Capítulo 9.

²²⁴ Yves Marcoux and Martin Sévigny. Why SGML? Why Now? *Journal of the American Society for Information Science*, 1997, vol. 7, nº 48, p. 584.

CAPÍTULO 6:

NORMALIZACIÓN

En la Sociedad de la Información a la que nos enfrentamos, donde se pretende un acceso global al conocimiento humano, el establecimiento de guías, normas y procedimientos estandarizados es fundamental para hacer los sistemas compatibles y facilitar la difusión y el intercambio automatizado de datos e informaciones a través de las redes de comunicaciones, especialmente en Internet. De todos es sabido que Internet es la Red de redes, global y no regulada; por ello, la necesidad de contar con estándares y con métodos consistentes para definir la información se ha convertido en un aspecto crucial para el propio progreso de la Red. Así, todos los programas de fomento de la Sociedad de la Información Global incentivan el desarrollo de estándares. Verbigracia, el programa de la Comisión Europea IST¹ (*Information Society Technology*) para conseguir una *Sociedad de la Información fácil de usar*, dentro del V Programa marco de IDT de la Unión Europea, destaca entre sus objetivos la necesidad de actividades de normalización y la creación de consenso para la innovación en el conocimiento.

Toda una incipiente familia de normas, especificaciones y recomendaciones de dominio público y/o estándares *de facto* (HTML, XML, RDF, etc.) están propiciando el cambio en la manera de prestar servicios y de acceder al contenido de la información electrónica. A lo largo del capítulo anterior, donde hablábamos de los modelos semánticos de metainformación (DC, TEIH, etc.) asociados a una sintaxis

¹ IST (*Information Society Technologies*): <http://www.cordis.lu/ist/overv-1.htm#objective>

(HTML/XML) o a una estructura (RDF), hemos intercambiado con bastante versatilidad los términos: "modelos", "formato", "esquema" y "estándar" de metadatos², para referirnos a distintas iniciativas de definición de elementos, estructuras y uso, destinadas a la organización de la Web. Sin embargo, el objetivo de este capítulo es reflexionar sobre la validez de estos modelos, esquemas o formatos y sobre cómo debe afrontarse el problema de la normalización en el entorno informativo de la World Wide Web para el que están constituidos. Para ello, trataremos la noción de normalización documental, mencionando la polémica surgida en torno a la validez y/o necesidad de los estándares en el ámbito de los metadatos; reflexionaremos también sobre los diversos casos de normalización, así como sobre el proceso de establecimiento de normas formales para los esquemas o modelos de metadatos tomando como ejemplo el caso del conjunto de elementos del Dublin Core; finalmente, plantearemos brevemente el complejo problema de la normalización del vocabulario de los metadatos refiriéndonos tácitamente a los metadatos basados en las materias, así como a los *schemes* de metainformación surgidos para completar los metadatos basados en el recurso.

A lo largo de este capítulo queremos destacar que la apreciación de la World Wide Web como un sistema global de información se funda en la normalización de la forma que adopta la información, no de las aplicaciones. En el ámbito general de Internet, la estandarización de los lenguajes de marcado abiertos, no propietarios y uniformes, es el fundamento de aplicaciones informáticas compatibles con cualquier plataforma, pudiendo intercambiar y visualizar la información independientemente de dichas plataformas. En el ámbito de la recuperación de información general en la WWW³, el desarrollo de sistemas de búsqueda genéricos obvia hoy en día la

² Como se desprende del capítulo 5, las estructuras, modelos, formatos y/o definiciones de metadatos se entienden con frecuencia, y de manera ideal, como estándares. Para una reflexión más profunda y sistemática sobre los términos asociados a la estructura y semántica de la metainformación. *Vid. Glosario, Metadata [Format/Model/Registry/Schema/Standard].*

³ *Vid. Capítulo 7.*

existencia de estándares de metainformación, lo que potencia nuestra tesis, también desde la normalización, de la aplicabilidad de metadatos en la constitución de sistemas de recuperación de información *ad hoc* en entornos informativos locales, aunque confiamos a la implantación efectiva de estos estándares y de su interoperabilidad, la eficacia de la recuperación global.

6.1. Metadatos: el problema de la estandarización

Desde la perspectiva documental tradicional (basada en estándares sólidos de organización del conocimiento —*v. gr.* las AACR2 o la LCSH—), la falta de normas compartidas y la inconsistencia de la indización de la información digital son los principales obstáculos para un acceso ilimitado y coherente a la información de la Web. Los grandes catálogos unificados, en los que las bibliotecas han confiado durante tanto tiempo para compartir e intercambiar registros, proyectos cooperativos, etc., que han permitido el acceso a grandes colecciones, habrían sido imposibles si no se hubiese empezado la normalización de formatos bibliotecarios en los años 60's. Cuarenta años después del inicio de la normalización MARC, las bibliotecas del siglo XXI están ante el mismo imperativo que protagonizó la normalización de la catalogación, en un mundo mucho más complejo y diverso de información electrónica.

Una norma, en general, siguiendo la Guía de la ISO/IEC⁴ es *un documento, establecido por consenso y aprobado por una institución reconocida, que proporciona, para el uso generalizado y común, reglas, guías o características para las actividades o sus resultados, que tiene como objetivo conseguir el grado óptimo de orden, en un contexto determinado*. Desde el punto de vista de la Documentación y del tratamiento de la información, las normas son pautas o directrices aceptadas por

⁴ Citado por la WSSN. World Standards Service Network. *General Info on Standardization* [documento HTML]. WSSN, 15 de noviembre de 1999. Disponible en: http://www.wssn.net/WSSN/gen_inf.htm #Standards (consultado el 25 de abril de 2001).

la comunidad, destinadas a promover un almacenamiento consistente de datos, de tal forma que se convierten en piezas indispensables para el intercambio de información eficiente⁵ y para una búsqueda y recuperación de información útiles. Así, Pat Oddy⁶ define los estándares bibliográficos como:

Aquellas normas que, cuando se aplican a una colección de información almacenada, organizan y facilitan el acceso a los registros de conocimiento y a su contenido intelectual.

En las bibliotecas y servicios de información se usan normas por varios motivos obvios: para incrementar la calidad y la estabilidad de la información, para mejorar la compatibilidad de estructuras de datos y para facilitar tanto la recuperación como el intercambio de información. Según esa definición y uso de los estándares, se puede decir que los metadatos deben de estar normalizados (son, o deben ser estándares), teniendo en cuenta que la utilidad principal presuntamente asociada a ellos —aumentar la relevancia en la recuperación e incrementar la consistencia de la información en un sistema de información electrónica distribuido como Internet— es la misma que se ha asociado tradicionalmente a las normas en Documentación. Si trasladamos la definición de Oddy al contexto de la información en Internet, podemos asimilar aún más el valor estimado de los metadatos, al valor de los estándares, ya que aquellos también se aplican a una colección⁷ de conocimiento (en este caso, "virtual" o digital y heterogénea) y sirven para facilitar el acceso a los registros de conocimiento y a su contenido intelectual.

⁵ Sobre este aspecto del intercambio de información mediatizado por la normalización, inciden muchos autores, entre ellos Pat Oddy, que dedica un capítulo — 4. "Normas para compartir"— a este tema, en su obra: *Future Libraries: Future Catalogues*. Esta autora comenta expresamente que los estándares son la base sobre la que las bibliotecas y unidades de información apoyan la posibilidad de compartir el contenido de sus colecciones de registros de conocimiento e información. Pat Oddy. *Op. cit.*, p. 58-77.

⁶ Pat Oddy. *Ibid.*, p. 58

⁷ Sobre el concepto de colección en el marco de las nuevas bibliotecas digitales, profundizaremos en el capítulo 8, sobre todo, al hablar de "colecciones virtuales" en las *subject gateways* (Cfr. 8.2.2).

A pesar de la lógica de estos planteamientos, partiendo del concepto general de norma y de la tradición normativa bibliotecaria matizada por la realidad de la información digital, existen posturas antagónicas en cuanto a la normalización de los metadatos. Algunos autores, llevan al extremo la concepción del carácter estandarizado de la metainformación, dudando, como Gorman⁸, del componente normativo de los metadatos, antes incluso de plantear la oportunidad de que sean estándares. Por ello, debemos dar cuenta aquí del debate en torno a la necesidad, oportunidad e incluso, conveniencia de hacer que los modelos de metadatos se conviertan en estándares.

Si profundizamos más en la normalización documental, podemos distinguir varios tipos de estándares, presentes en la tradición de la gestión de la información:

- a) Normas en torno a la estructura. Esto es, aquellas que definen los campos y sus relaciones (como el formato MARC).
- b) Estándares sobre el contenido, donde incluiríamos tanto las reglas de catalogación como la sintaxis de los datos.
- c) Y finalmente, normas destinadas a la valoración semántica de los datos, es decir, vocabularios específicos que definen la forma y la importancia del contenido.

En términos semejantes a nuestro planteamiento, Tony Gill⁹ habla sobre la aplicación de normas a los sistemas distribuidos de información, identificando cuatro categorías de estándares que deben considerarse: *técnicos, de estructura, de contenido y de organización. Las siguientes normas pueden considerarse en algunas*

⁸ Michael Gorman. From Card Catalogues... *Op. cit.*, http://www.unicamp.br/bc/gorman_paper.html

⁹ Tony Gill en una conferencia celebrada en 1997 (*Bournemouth University Library Information Systems Conference'97*). Citado en: John Eyre. Distributed Image Services (Working Together). *Vine (Theme issue: Digital Images in Libraries)*, July 1998, n° 107, p. 67.

de estas áreas: *TCP/IP*, *Z 39.50*, *Whois++*, *AACR*, *MARC*, *CIDOC*, *AAT*, *Dewey Decimal*, *Spectrum* y *el modelo de metadatos en desarrollo del Dublin Core*. Lo que hace Gill, no es más que exportar la tipología tradicional de normas documentales al nuevo contexto de información distribuida, agregando además un tipo indispensable al hablar de entornos automatizados: los estándares técnicos como los protocolos *TCP/IP* ó *Z39.50*. Después del reconocimiento de estos cuatro tipos de normas, el autor enumera varios de los estándares utilizados en Internet pero sin clasificarlos dentro de ningún tipo concreto, lo que denota el reconocimiento de su uso, pero el carácter inclasificado, o de difícil clasificación, de algunos de ellos (por ejemplo *Z39.50* se puede considerar un estándar técnico, pero también de estructura). Así, reconoce (en 1997) la proyección del DC como estándar de metadatos cuya evolución había comenzado en 1995; sin embargo, no lo clasifica dentro de ningún tipo, ya que un formato de metadatos, desde nuestra visión y como hemos podido dilucidar en la descripción de algunos de ellos en el capítulo 5, comporta todo un entorno normativo de estructura, sintaxis, semántica y contenido.

De igual forma que la organización de la información convencional está repleta de normas para la recuperación y el intercambio en bibliotecas y servicios de información, en la organización de la información digital (para promover y facilitar el intercambio y recuperación de información en bibliotecas digitales y servicios de información electrónica), la representatividad y necesidad de estándares es aún más ostensible. Los metadatos son o deben ser, como apuntamos aquí, normas relacionadas con otras normas, bien de carácter bibliográfico (como el *MARC*), bien de tipo estructural o sintáctico (*XML/RDF* u otras como *GRS-1*, *ISO 2709* o *SUTRS*), y relacionadas también con protocolos de intercambio (desde *HTTP* a *Z39.50*, pasando por *WSDL*, *SOAP*), etc. En los metadatos confluyen pues, todos los aspectos relacionados con la normalización documental y todos aquellos relacionados con estándares técnicos de la infraestructura de comunicaciones de la Red. Esta afirmación que reflejamos aquí a tenor de la normalización justifica tácitamente toda la complejidad que se advierte en los metadatos, ya que en ellos están implícitas

todas las cuestiones relacionadas con la gestión bibliotecaria y todas aquellas relacionadas con las más modernas tecnologías de la información.

De la ratificación implícita de Gill sobre la normalización de los metadatos en los sistemas distribuidos de información, unida a su finalidad equiparable a la de las normas documentales (compartir y recuperar información), se infiere que los modelos de metadatos están regidos por estándares, o al menos, que deberían estarlo. No obstante, nos parece oportuno evaluar ahora esta afirmación fácilmente deducible de lo anterior, intentando responder a las siguientes cuestiones: ¿En qué medida se deben normalizar los modelos de metadatos, de modo que puedan ser la base para constituir catálogos distribuidos y unificados de registros de conocimiento en formato electrónico, recuperables y compartibles? ¿Por qué es importante contar con estándares de metadatos? ¿Qué beneficios, riesgos y oportunidades implica la normalización? ¿Qué tipos de estándares se relacionan con los metadatos? ¿Cuáles serían los requisitos de un estándar de metadatos?

A lo largo de esta investigación hemos apuntado distintos aspectos de los metadatos con el fin de ilustrar su necesidad de normalización, para que permitan mejorar la recuperación de información. En aras a una mayor claridad de lo que se va a exponer en este apartado, vamos a sistematizar algunas cuestiones y/o premisas en torno a la metainformación, que debemos subrayar en virtud a su normalización:

- El control del vasto número de recursos existentes en Internet, requiere un "formato de registro"¹⁰ que permita una descripción oportuna de los recursos, así como su posterior localización o recuperación.
- Los metadatos deben contemplar todos los elementos que permitan al usuario examinar la relevancia potencial de un documento sin la necesidad de visualizar el objeto (DLO) completo.

¹⁰ Cfr. Glosario, *Metadata [Format/Model/Registry/Schema/Standard]*.

- Los registros deben ser compatibles con un motor de búsqueda, que a su vez sea compatible con un protocolo de búsqueda y recuperación de información en Internet (v. gr. Z39.50, WHOIS++); además, es preciso que todos estos componentes se conformen, simplifiquen o describan de forma estandarizada y a nivel internacional.
- Como hemos visto en el capítulo dedicado a los formatos, un solo estándar o pequeño grupo de estándares no será capaz de realizar esta tarea, por lo que debe de existir una interoperabilidad entre los modelos de metadatos desarrollados por las distintas disciplinas o entornos de información, y las últimas tendencias (Warwick Framework/RDF) así lo demuestran. En otras palabras, no es tan necesario seguir un único estándar, como crear un marco normativo genérico que permita la compatibilidad de diferentes modelos de metadatos.
- Hemos hecho alusión también a la necesidad de que los metadatos no sean sólo iniciativas caprichosas de distintas áreas científicas (metadatos con propósitos específicos, para la información educativa, geoespacial, gubernamental, etc.), que respondan a las necesidades de recuperación de información puntuales dentro de cada área, sino que constituyan estándares reconocidos y compartidos para el intercambio de información.

Según estas conclusiones derivadas del propio concepto de metadatos, de su utilidad para la recuperación de información, de su proceso de asignación y de los distintos formatos surgidos para cubrir esa necesidad, parece obvio que son necesarias normas sólidas que regulen la descripción del conocimiento albergado en la Red, cuyas ventajas para recuperar y compartir información son innegables. La importancia de la normalización en el ámbito más genérico de la información electrónica —en el que también se incluyen los metadatos— ha sido reconocida por distintos autores e instituciones. Por ejemplo, Rodríguez Muñoz¹¹ habla del

¹¹José Vicente Rodríguez Muñoz. Documentos electrónicos y normalización: información y conocimiento. Perspectivas de futuro. *Scire*, enero-junio 1997, vol. 3, nº 1, p. 139.

establecimiento de normas, de metodologías y de formas de hacer que sean ampliamente aceptadas y asumidas, como una *cuestión trascendente* para alcanzar un grado de *igualdad en el acceso al conocimiento*. Reflexiona también sobre el nivel al que hay que llevar la normalización: no hasta el *extremo del ahogo normativo*, pero sí propone, al menos, *programar estrategias, recomendaciones, dictámenes, informes, estándares, etc.*¹². Este nivel de "recomendaciones" o estándares *de facto* (no de ahogo normativo) es el que nos parece más adecuado para la metainformación. No obstante, consideramos necesario evaluar no sólo las ventajas aparejadas a la normalización, sino también los inconvenientes y/o dificultades de constituir estándares de metadatos.

En cuanto a los beneficios que se pueden obtener de la aplicación de normas de metainformación, podemos destacar los siguientes:

- Facilitan el mantenimiento y actualización de los recursos, tanto dentro de un contexto informativo finito o local (biblioteca digital) como en toda la Web.
- En contextos genéricos de información, esto es, metadatos de propósito general destinados a optimizar la recuperación en toda la Web, la normalización implicará una aplicación consistente de metaetiquetas. Hemos destacado, sobre todo al hablar de las estructuras HTML para la metainformación (5.2.1.1) y de las herramientas para la creación de metaetiquetas (4.3.1), y volveremos a mencionar este tema al comparar los sistemas de búsqueda en texto completo frente a los sistemas de recuperación de información basados en metadatos¹³, la veleidad o inconsistencia de los creadores de información al asignar metaetiquetas. El hecho de que existan normas y manuales de procedimiento en la aplicación de

¹² *Ibid.*, p. 143. Sobre el "ahogo" que pueden suponer las normas en la gestión de la información, *Cfr.* Pat Oddy, *Op. cit.*, p. 75.

¹³ *Cfr.* 7.3.

metadatos, asegurará una aplicación más coherente de ellas, así como la consiguiente absorción de la norma por los agentes de software.

- En contextos finitos de información, los documentos estructurados conforme a un estándar de metadatos son más fáciles de gestionar y pueden dar lugar a la formación coherente de una biblioteca digital y del sistema de recuperación asociado a ella.
- Aseguran una búsqueda de información más precisa y relevante, fundada en el propio esfuerzo o política de normalización y en la coherencia que ésta impone al sistema.
- Permiten la gestión automática de los sitios Web, como por ejemplo la creación de mapas de conocimiento¹⁴.
- La realización y adopción de estándares de metadatos entre los distintos autores de información digital así como entre los equipos creadores de colecciones digitales, permitirá la creación de sistemas uniformes y globales.
- En la integración de contenidos en toda la Web, bien de DLOs creados en un contexto finito de información (biblioteca digital¹⁵), bien de documentos creados sin ninguna finalidad particular, si existen estándares sólidos internacionales, el contenido se organiza y describe *a priori*, antes de la publicación de los documentos —bien por los autores o por los editores digitales en la formación de

¹⁴ Ya hemos destacado que algunas de las aplicaciones de creación y gestión de metadatos, como por ejemplo *Metamanage* permitan la generación de mapas del sitio a partir de estructuras bien definidas de metadatos. (Cfr.4.3.2.2).

¹⁵ Cfr. 8.2.1.

colecciones—; esto restará el poder de conjeturar metainformación *a posteriori* a las herramientas de indización automática destinadas a toda la Web¹⁶.

Una ventaja innegable de cualquier tipo de norma, de un estándar de metadatos también, es que se establecen a través de un proceso consultivo (con otros "expertos") y proporcionan la base desde la que se desarrollarán perfiles a nivel nacional, internacional u orientados a una disciplina o temática particular. Así, si un estándar se acepta dentro de una comunidad amplia, las herramientas informáticas se desarrollarán (o deberían desarrollarse) para la aplicación de esa norma a nivel industrial.

Los metadatos requieren una consistencia en su contenido y estilo para asegurar el intercambio de información. Teniendo en cuenta esto, debemos reflexionar si es preciso que los modelos de metadatos sean estándares formales, esto es, si es esencial o no implantar mecanismos escrupulosos y colectivos para el mantenimiento de normas de metadatos. En este sentido, en el encuentro sobre metadatos desarrollado

¹⁶ En este sentido, queremos destacar los dos ámbitos de aplicación que se recogen en esta tesis doctoral en cuanto a la aplicación de metadatos para la recuperación de información. Por un lado, el contexto general de producción y recuperación de páginas para la WWW (que reflejaremos en el capítulo siguiente, 7) y por otro, las bibliotecas digitales (que plantearemos en el capítulo 8). En ambos casos, contar con estándares para la recuperación de información puede ser beneficioso: en el primero, para incrementar las posibilidades de recuperación de los servicios genéricos instando a los autores de páginas Web a la inclusión de metaetiquetas normalizadas; y en el segundo, porque permitirá la creación de sistemas de recuperación *ad hoc* para un conjunto finito de documentos (biblioteca digital) o en la formación de colecciones virtuales (*subject gateways*).

Destacamos en este punto, sin embargo, que ambas filosofías de creación de documentos pueden destinarse a la recuperación global, es decir, el hecho de que se formen colecciones y sistemas de recuperación *ad hoc* sobre éstas en el ámbito de las bibliotecas digitales y de las *subject gateways*, no es óbice para que los documentos que pertenecen a una colección particular se recuperen de una forma global. Este es el principio que subyace a la formación de la *Web semántica*; sin embargo, en este momento, la recuperación genérica se asegura sólo a través de motores de búsqueda. Así lo reflejaremos en el diseño de aplicación de metadatos a la BVC del capítulo 9: por un lado metadatos destinados a la recuperación finita dentro de la propia biblioteca digital (esquema RDF y perfil de aplicación del DC adaptado a las necesidades de la BVC) y por otro, la proyección de esos metadatos para que sean tenidos en cuenta por los buscadores globales de la Red (etiquetas <META> de HTML).

en el ámbito del *Research Library Group* (RLG)¹⁷ en 1997, se apuntaban los siguientes alegatos matizando el concepto de normalización en el ámbito de la metainformación, que resumimos e interpretamos a continuación:

- Antes de que se imponga un proceso formal de normalización, se necesita una masa crítica de experiencias reales en la aplicación de metadatos. Ésta fue una de las reprobaciones realizadas, en un principio, al formato del Dublin Core como estándar de metadatos para la recuperación de información en Internet¹⁸.
- Las diversas comunidades de investigación deben centrarse en alentar sus propios esfuerzos de defensa y formación. Cada comunidad tiene ya sus propios mecanismos para conseguir acuerdos y para perfeccionar los atributos de metadatos dentro de su propio dominio de aplicación.

El estado de los metadatos ha mostrado durante mucho tiempo ya, que cada comunidad está volcada a defender sus necesidades de información así como los modelos de metadatos que las reflejen. A partir de 1999, y como hemos apuntado, a raíz de la recomendación del modelo y la sintaxis de RDF, se empieza a generar una mayor conciencia de integración de formatos y una necesidad creciente de normalización formal.

- El proceso de estandarización formal, tal y como está establecido, evoluciona demasiado lentamente para ajustarse a las necesidades del entorno de información en Internet.

¹⁷ Willy Cromwell-Kessler and Ricky Erway. *Metadata Summit (1997. Mountain View): Meeting Report* [documento HTML]. Research Library Group, rev. 7 de julio de 1997. Disponible en: <http://www.rlg.org/meta9707.html> (consultado el 24 de julio de 1998).

¹⁸ Recordemos que en octubre del mismo año de la conferencia del RLG (1997) se celebró en Helsinki el quinto *workshop* del Dublin Core (*Cfr.* 5.2.1.2.), donde ya se entendía al DC como un formato estable que requería una normalización formal. Posteriormente, como veremos, se irá matizando esta afirmación sobre la estabilidad del DC. *Cfr.* 6.3.

De todos es sabida la demora de los procesos de normalización, la burocracia y revisión de borradores que conlleva, que pueden hacer que un proceso de estandarización dure incluso años. Éste quizás sea uno de los aspectos clave, un inconveniente trascendental en el establecimiento de normas formales de metadatos, ya que la evolución de Internet es tan vertiginosa que hace prever que los estándares formales pueden quedar obsoletos incluso antes de llegar a aprobarse, como ocurrió con la norma ISO relativa a HTML¹⁹.

A todos los riesgos que señalaba el Grupo de Investigación Bibliotecaria (RLG) y que le llevaban a concluir en 1997, que era demasiado pronto para intentar procesos de estandarización formal, podemos añadir otras afirmaciones generales, no vinculadas a una apreciación temporal, que constituyen también riesgos en el sometimiento de modelos de metadatos a procesos de normalización:

- La práctica bibliotecaria tradicional está repleta de metadatos —en sentido amplio— estricta y formalmente normalizados a nivel internacional (ISBD, UNIMARC, etc.) y nacional (AACR, USMARC, IBERMARC, etc.). En muchos casos, esta excesiva estandarización dificulta, retarda e incluso detiene el proceso de gestión eficaz de la información.
- A tenor de experiencias normativas en el mundo documental tradicional, como es el caso de las reglas de catalogación o el formato MARC (que, como dijimos, podría considerarse un modelo de metadatos²⁰), es preciso señalar que la más

¹⁹ ISO/IEC 15445:2000 / DCOR 1:2001(E). *Information Technology - Document Description and Processing Languages - HyperText Markup Language (HTML)* [documento HTML]. Dublin: Trinity College, Computer Science Department, 15 de mayo de 2000. Disponible en: <http://www.cs.tcd.ie/15445/15445.html> (consultado el 18 de agosto de 2000).

²⁰ Murtha Baca dice que el MARC puede verse como un tipo de "contenedor" técnico —norma cum metadata—, es decir, un conjunto específico de campos y subcampos designados para contener tipos específicos de valores de datos, con reglas específicas de aplicación. Murtha Baca. *A Picture Is Worth a Thousand Words: Metadata for Art Objects and Their Visual Surrogates*. 2000, p. 1, (ejemplar cedido por de autor). Esta visión de Baca acerca del MARC, matiza de forma atinada a nuestro juicio, algunas de las afirmaciones que trepidan en la bibliografía sobre este tema, que

[cont.]

rigurosa aplicación de estándares no garantiza siempre la consistencia de la información.

- Por otra parte, si los distintos formatos fuesen normas inflexibles y complejas, se comprometería la necesidad de personal especializado para la aplicación de las mismas se dificultaría sensiblemente el principio de creación por parte del autor/editor del documento Web²¹.

Esta reflexión sobre los riesgos de aplicar normas formales a los metadatos responde tácitamente a lo que Pat Oddy²² argumenta categóricamente bajo el marbete de *normas de piedra*. Esta autora relaciona los estándares con el desarrollo tecnológico, basándose en la asunción de que el cambio tecnológico invalida el propósito general para el que se han desarrollado tradicionalmente los estándares y considera que:

Las normas ahogan la innovación y cuanto más extienden sus tentáculos y más penetrantes se hacen, más difícil es pensar en las funciones que proporcionan a la luz de las nuevas tecnologías [...]

Internet supone el mayor cambio tecnológico de la historia de la producción de conocimiento y, por ello, de la historia del almacenamiento y recuperación de información —Oddy, refiriéndose al cambio tecnológico continúa— [...] *pero es poco probable que cambie el propósito fundamental de los bibliotecarios de adquirir, describir y hacer accesible el conocimiento intelectual de sus colecciones. No obstante, debemos ser conscientes de que la forma de acceso a ese conocimiento*

entienden categóricamente el MARC como un modelo más de metainformación. El formato MARC es pues un *estándar cum metadata*, una norma que funciona al estilo de los modelos de metadatos para la información electrónica pero sin ser considerado expresamente uno de ellos. *Cfr.* 4.1.

²¹ Es preciso reconocer que esto que aquí apuntamos como riesgo de una normalización férrea y formal de los modelos de metadatos —frente al principio comúnmente aceptado (*Vid.* 4.2.) de creación e metadatos por los autores del documento— puede convertirse en una oportunidad para la redefinición del papel de los profesionales de la información ante las nuevas bibliotecas digitales y los nuevos servicios de información.

²² Pat Oddy, *Op. cit.*, p. 75. *Cfr.* José Vicente Rodríguez Muñoz. *Loc. cit.*

*intelectual es, en cierto modo, diferente según el medio en que se encuentre almacenado y según las necesidades particulares de los usuarios*²³.

En cualquier caso, y a pesar de que permanezca la filosofía primigenia de facilitar el acceso a la información, es inevitable que, tal y como se plantea la perspectiva de la asignación de metadatos, podamos pensar en una redefinición del papel profesional, paralela al cambio que supone la descripción de objetos de información (DLOs). Por ello, es preciso evaluar la eficacia concreta que tendrían unas normas férreas aplicadas a los metadatos en el nuevo entorno tecnológico de la Red donde, evidentemente, cambia tanto la forma de acceso como las necesidades de los usuarios.

Desde nuestro punto de vista, aún es pronto para evaluar la eficacia de normas de este tipo por tres motivos fundamentales: por un lado, porque consideramos que aún es pronto para evaluar la eficacia de los estándares de metadatos en general, por otro, porque ninguno de los modelos de metainformación puede ser considerado una norma férrea y, finalmente, porque se percibe una tendencia a flexibilizar la aplicación de estándares, al adecuarlos a las necesidades concretas de cada comunidad informativa de la Web (los *application profiles* del DC²⁴, son una muestra evidente de esta visión).

²³ Pat Oddy. *Ibid.* Esta idea de Oddy va en consonancia con la filosofía global que queremos infundir en esta tesis: las tecnologías de la información cambiarán *el cómo* del trabajo bibliotecario no *el por qué*, por contraposición a la tendencia, demasiado vinculada a la gestión de la información tradicional impresa, que estima que la Red no va a cambiar el *modus operandi* del bibliotecario, lo que conlleva perpetuar las tareas tradicionales, como la catalogación en sentido estricto, en Internet.

Disentimos con esta última visión del cambio tecnológico, concretamente, en lo relativo a la innovación que supondrá Internet, ya que se modifica, en principio, el concepto de colección y el concepto de ítem de información (objetos de información, DLOs que pueden ser desde un sonido, una página o un conjunto de ellas, un sitio Web, etc. hasta una imagen o un vídeo). Tampoco podemos hablar de adquisición en sentido estricto, como un *propósito fundamental* del trabajo bibliotecario. Vid. 8.1, 8.2. y 9.2.2.

²⁴ Vid. 5.2.1.2.

Por otra parte, una ausencia de normalización total llevaría a una organización anárquica en la gestión de la información, que redundaría obviamente en la recuperación, donde sólo se confiaría la relevancia a sofisticados algoritmos de búsqueda de los software de indización automatizada. No en vano, esta bibliotecaria británica (Oddy), en la misma obra que venimos citando (*Future Library: Future Catalogues*) pone el ejemplo de los usuarios de Internet que visitan distintos sitios de transferencia de ficheros (FTP), para ilustrar que la carencia total de normalización repercute en la recuperación y las dificultades que supone la búsqueda en bases de datos que no se han creado atendiendo a un estándar común:

[...] los fondos de información FTP están organizados únicamente de acuerdo con los deseos de la persona que establece el archivo. Encontrar ítems puede ser complicado, y conocer qué ficheros estás hojeando a partir de su contenido real puede ser difícil, teniendo en cuenta que los directorios sólo listan el nombre del fichero y el tamaño en bytes del mismo [...] y algunos sitios, lamentablemente no tienen ningún fichero de índice²⁵.

Esta digresión acerca de la transferencia de ficheros es un ejemplo evidente de la necesidad de una normalización y una descripción de los recursos de Internet, máxime cuando muchas páginas Web comportan, entre otros DLOs, ficheros de datos o programas que se pueden descargar²⁶.

Según todo lo que hemos apuntado sobre la necesidad o no de normalización, nos decantamos por una solución mixta para los metadatos: ni la normalización exacerbada o el *ahogo normativo*²⁷, *el colmo de la complejidad* —como dice

²⁵ Evoca el consejo aparecido en la revista de la empresa *CompuServe*: *Discovering FTP. CompuServe Magazine*, March 1998, p. 28. Citado en: Pat Oddy. *Op. cit.*, p. 62.

²⁶ Recordemos que los DLOs para cuya descripción se pueden utilizar metadatos (más o menos normalizados), pueden ser también un programa de ordenador y que uno de los primeros formatos de metainformación (IAFA) estaba asociado al protocolo FTP.

²⁷ José Vicente Rodríguez Muñoz. *Loc. cit.*

Leloup²⁸—, del mundo bibliotecario en la era del *Jurassic MARC*²⁹, ni la falta total de normas de estructuración y contenido semántico de la información. Se debería aprovechar la agilidad que supone el asignar metadatos sin unas normas rígidas, y hacer que los formatos de metadatos fuesen, más que normas formales, "modelos normalizados" asumidos por una comunidad de autores, editores, informáticos, usuarios, y bibliotecarios. Además, el propio desarrollo y evolución de los lenguajes de descripción de documentos o de marcado, permiten que los modelos normalizados de metainformación descansen a su vez en otros estándares *de facto* flexibles y escalables como el metalenguaje XML o el "metamodelo de metadatos" RDF.

Como reconoce Richard Gartner³⁰, las bibliotecas hoy en día conciben el mundo digital como parte u objeto de su misión de crear bases de datos³¹, a menudo acompañando versiones digitales de sus fondos, pero también como servicios aparte. Por ello, *sin ningún método de normalización, especialmente en el área de formatos de información, será difícil relacionar estas bases de datos con otras*, ni dentro de una misma biblioteca, ni mucho menos con otras bases de datos específicas

²⁸ Catherine Leloup, hablando de la representación de la información dice que *el colmo de la complejidad se obtiene con los formatos de reseñas bibliográficas, los formatos MARC como Machine Readable Cataloging, de los cuales las malas lenguas afirman que la simple lectura de sus cientos de campos, simples o múltiples, con subcampos y más subcampos deja extenuados a los ordenadores*. Catherine Leloup. *Op. cit.*, 41.

²⁹ Esta ingeniosa y sutil metáfora aprovechando los parónimos en inglés "PARK" y "MARC", la utiliza Pat Oddy —*Op. cit.*, p. 63— para referirse al entorno total y formalmente normalizado en documentación, por contraposición a la anarquía, desde el punto de vista documental, que supone la organización de la información en Internet.

Hudgins, Agnew y Brown, se refieren al exceso de normalización bibliotecaria clasificando al MARC como un *estándar monolítico*. Concretamente, aluden a que las tecnologías y las normas emergentes permitirán a los distintos formatos de metadatos interoperar e incluso coexistir en la misma base de datos de una forma más fácil y flexible, que la búsqueda en diferentes colecciones, que ha permitido el monolítico MARC hasta ahora. Jean Hudgins, Grace Agnew, Elizabeth Brown. *Op. cit.*, p. 60.

³⁰ Richard Gartner. *Op. cit.*, p. 158.

³¹ De ahí que en el entorno bibliotecario se mantengan proyectos como CORC. *Cfr.* 4.1.

existentes en la Red. Por ejemplo, la consulta a catálogos Z39.50 que ha potenciado los servicios de las bibliotecas conectadas a Internet —tanto desde el punto de vista del trabajo técnico (captura de registros catalográficos) como del servicio a los usuarios (consultas uniformes a distintos catálogos)—, es posible gracias al estándar ISO 2709 para el intercambio de información. La consistencia en el contenido y estilo de los metadatos es necesaria para asegurar el intercambio de información.

Desde ese punto de vista, parece inevitable que los formatos de metadatos se conviertan en normas. Por ello, ahora la pregunta es ¿qué debería comportar un estándar de metadatos? En este sentido, evocamos la definición de carácter topológico (ya que no se refiere a ningún formato en particular) que da Linda Hill³² en el taller sobre metadatos de la ADL, donde afirma que los estándares de metadatos son un desglose de los siguientes componentes:

- *Un objetivo o propósito: el intercambio de información*, que hace que los estándares de metainformación sean la guía donde se recoja una regulación efectiva para la creación de descripciones.
- *Elementos de datos*: es decir, ítems de información a los que va dirigido. En este sentido es importante que dichas normas contemplen el tipo de elementos informativos para los que se establecen.
- *Estructura*: etiquetas, jerarquía o anidamiento³³. Un estándar de metadatos deberá definir también el conjunto de etiquetas así como las técnicas necesarias para su relación.

³² Esta idea se recoge en el informe de Michael Goodchild en la *Alexandria Digital Library Workshop on Metadata*. Michael F Goodchild. *Op. cit.*, http://www.alexandria.ucsb.edu/public-documents/metadata/metadata_ws.html

³³ *Vid.* Glosario, *nesting*.

- *Guías para la consignación de los datos.* Las normas deben interrelacionar un conjunto de formatos, fuentes de información, etc. que hagan factible su aplicación.
- Y por último, según Hill, deben prever también *valores de datos controlados*, esto es, listas de encabezamientos de materia, tesauros, listas de categorías, etc.

De todos estos elementos a los que se refiere esta bibliotecaria de la ADL, y teniendo en cuenta que los estándares de metadatos no deben constituirse sólo para dotar de consistencia interna a una colección particular de DLOs (v. gr. imágenes de arte, documentos Web geoespaciales, etc.) sino que deben conseguir una consistencia extendida a otros sistemas, se deduce que compartir información es la finalidad principal de la normalización, y debe serlo también, en los entornos estandarizados de metadatos. Sin embargo, para dar coherencia a sistemas de información finitos basados en metadatos, es preciso contar con manuales de procedimiento acordes al sistema de recuperación local.

Con todas las reflexiones que hemos reflejado aquí, desde nuestro punto de vista, los sistemas de información distribuida generados en Internet deben fundamentarse en entornos normalizados de metadatos abocados a la interoperabilidad, más que en estándares formales de metadatos. Esos metadatos normalizados *de facto* deben tener una estrecha relación con los estándares técnicos como XML, ya que tiene el potencial de funcionar como norma aglutinante para constituir lenguajes de marcas específicos para cada tipo de documento y para fundar sistemas de información flexibles e integrados técnicamente; o con estándares diseñados expresamente para la recuperación de información, como ANSI/NISO Z39.50³⁴, basado en la arquitectura cliente/servidor diseñado para soportar la

³⁴ Vid. *A Guide to the ANSI/NISO Z39.50 Protocol*. Bethesda, U.S.: National Information Standards Organization, 1995, 12 p. *Z39.50 profile for access to digital collections*. Washington, DC.: Library of Congress, 3 de mayo de 1996, rev. 13 de enero de 1997. Disponible en: <http://lcweb.loc.gov/z3950/agency/profiles/collections.html> (consultado el 6 de julio de 1998). Ver también todas las modificaciones, atributos, etc. que se han propuesto para este protocolo después de la reunión del ZIG [cont.]

búsqueda y recuperación de información en un entorno de red distribuido como Internet. Por otra parte, además de esta relación con estándares técnicos, hay quienes, como por ejemplo el Comité de Catalogación de la ALA³⁵, establecen la necesidad de relación entre los metadatos, con otros estándares de contenido (como las AACR2) o de valor semántico (como los tesauros)³⁶.

En cualquier caso, es preciso controlar y normalizar tanto la sintaxis (XML/SGML o HTML), como la estructura (RDF) y la semántica (DC, TEI, etc.) de los datos legibles por ordenador, como la sintaxis (ISBD, AACR) y la semántica (vocabularios y listas de autoridad o materia), entendibles por el hombre. No obstante, a pesar de que los metadatos se relacionen con las normas formales de la Documentación, creemos que los nuevos sistemas de recuperación de información electrónica basados en ellos no deben depender de la rémora de las ISBD o de las clasificaciones bibliotecarias y que deben vincularse a la realidad flagrante de la Web Semántica.

6.2. Estándares de metadatos: casuística y formalización

A pesar de todo lo que hemos dicho acerca del tedio que supone la formalización de normas, a raíz del desarrollo y utilización de metadatos, algunos de estos modelos se han involucrado en procesos de estandarización formal, bien a través de las organizaciones nacionales e internacionales de normalización, como el

de octubre de 2001 en Boston, de la cual ha surgido un nuevo borrador de la norma americana. *ANSI/NISO Z39.50-2001. Op. cit.*, <http://www.loc.gov/z3950/agency/revision/part1.pdf> y <http://www.loc.gov/z3950/agency/revision/part2.pdf>

³⁵ American Library Association. *Committee on Cataloging: Description and Access Task Force on Metadata and Cataloging Rules: Final Report* [documento HTML]. ALA, 3 de junio de 1998, rev. 4 de junio de 1998. Disponible en: <http://www.ala.org/alcts/organization/ccs/ccda/tf-tei2.htm> (consultado el 18 de junio de 1998). En este informe de la ALA, se relacionan dos modelos de metadatos con las Reglas de Catalogación Anglo-Americanas (norma nacional americana de contenido basada en las ISBD).

³⁶ *Cfr.* 6.4.

ANSI (*American National Standard Institute*) o la ISO (*International Standard Organization*), bien a través de los nuevos organismos con entidad normalizadora en la Web, como el Consorcio (W3C) o el IETF (*Internet Engineering Task Force*), que anticipan, en muchos casos, la formalización de estándares. Por ello, en este apartado vamos a analizar el porqué del establecimiento de normas formales de metadatos, así como de la compleja casuística de la normalización de Internet que se complica aún más en la cantidad de formatos y modelos de metainformación existentes y de sus pretensiones, siempre normativas.

La primera circunstancia que debemos destacar es la capacidad normalizadora de instituciones como el Consorcio Web, que nos hace distinguir tres tipos de estándares en Internet a tenor de su estabilidad y/o aplicabilidad:

- a) Estándares *de facto*. Se producen cuando un producto se establece de tal forma que la compatibilidad se mide por referencia a él. Suelen tener un nivel de reconocimiento y aceptación generalizado en el mercado e imponen que cualquier desarrollo ulterior, los tenga en cuenta, como por ejemplo los formatos de Microsoft (.doc, .ppt, .xls, etc.) o, cada vez más, también los de Adobe (.pdf, por ejemplo). En el ámbito de los metadatos, este sería el caso, *v. gr.* de OpenURL, una tecnología desarrollada por la empresa de automatización de bibliotecas Ex-Libris³⁷, que ya ha sido adoptado, aunque todavía no se ha constituido como norma formal³⁸, por otras empresas y servicios de información como OCLC³⁹ o el grupo Gale⁴⁰.

³⁷ *Cfr.* Capítulo 4, notas 12 y 92.

³⁸ No obstante, como comentaremos a continuación, la tendencia es que los estándares *de facto* acaben normalizándose formalmente. En el caso de OpenURL <<http://www.sfxit.com/OpenURL>>, en junio de 2001 se iniciaron las primeras reuniones para convertir esta norma en estándar nacional ANSI/NISO. Más información en: <http://www.niso.org/commitax.html>

³⁹ OCLC anunció el 25/10/01 la adopción de este estándar *de facto* en su servicio FirstSearch (*Vid.* Capítulo 4, nota 35). OpenURL es un protocolo basado en Web para la transmisión de metadatos que permite aumentar la visibilidad y accesibilidad de los recursos a través de un sistema de enlaces que

[cont.]

- b) Estándares formales —o *de jure*⁴¹— esto es, aquellas normas formalmente establecidas por ley o por una institución reconocida para la formulación de estándares. Dentro de este grupo podemos clasificar los estándares de la ISO o de la ANSI (por ejemplo ISO 8879, ó ANSI/NISO Z39.50).
- c) Y también se distingue un tipo mixto de los dos anteriores que podríamos denominar con el nombre paradójico de "estándares formales de facto", formado por todos aquellos que, sin haber sido reconocidos por una institución con potestad normalizadora, han sido proclamados y/o recomendados por una institución con responsabilidad y participación internacional en el desarrollo de Internet.

Estas últimas normas que nosotros llamamos "formales de facto" son las que las *Guidelines on Best Practices for Using Electronic Information*⁴² de la Unión Europea, denominan Especificaciones de Dominio Público (PAS) y que definen como estándares desarrollados por algunas firmas líderes en el mercado que se unen en forma de consorcio para definir una norma común. El hecho de definir estándares de interconexión les permitirá a las distintas empresas desarrollar productos mutuamente compatibles. El caso más significativo de este tipo de consorcios es el del W3C que, liderado por Tim Berners-Lee, creador de la World Wide Web, y con Jean-François Abramatic como presidente, está constituido por

genera entre ellos. OCLC ha sido el primer servicio que ha reconocido la utilidad de esta norma para enlazar citas entre las bases de datos de FirstSearch y los textos completos de artículos en otros recursos proporcionados por las bibliotecas. *OCLC Offers OpenURL Option for Libraries* [documento HTML]. Dublin: OCLC, 25 de octubre de 2001. Disponible en: <http://www.oclc.org/oclc/press/20011025.shtm> (consultado el 19 de noviembre de 2001).

⁴⁰ *Gale Adopts Sfx Open Url Standard* [documento HTML]. Galegroup, Thompson Learnig, 31 de octubre de 2001. Disponible en: http://www.galegroup.com/press_room/article/oct01_sfx.htm (consultado el 19 de noviembre de 2001).

⁴¹ *Guidelines on Best Practices for Using Electronic Information: How to Deal with Machine-readable Data and Electronic Documents*. Updated and enlarged ed. Luxembourg: Office for Official Publications of the European Communities, 1997, p. 17, 50.

⁴² *Ibid.*, p. 50.

varias organizaciones sin ánimo de lucro que trabajan con la comunidad internacional para desarrollar especificaciones y programas informáticos gratuitos de referencia. Los servicios del W3C implican, al menos en principio: información sobre la *World Wide Web* para desarrolladores y usuarios, realización de códigos de referencia para incorporar y promover estándares y varias aplicaciones de demostración para validar el uso de nuevas tecnologías.

Los estándares *de facto* y las normas formales, se reconocen en cualquier ámbito susceptible de normalización. Sin embargo, el tercer tipo que hemos señalado (*Public Available Specifications* o normas formales de facto) está especialmente vinculado al mundo de la tecnología y, en particular, de la edición electrónica en la Web. Constituyen, desde nuestro punto de vista, una respuesta, por un lado, al reconocimiento de la necesidad de normalización en el ámbito de Internet y, por otro, a la dilación generalizada de los procesos formales de creación y publicación de estándares.

Además de estos tipos de normas, la versatilidad, universalidad y facilidad para el intercambio de información que implica la propia Web como medio de comunicación, hacen que exista, en el ámbito los metadatos, otro tipo de estándares o, más bien, otro estadio en la normalización, situado a un nivel previo a las especificaciones de acceso público pero con tanta visibilidad y aceptación como éstas. Nos referimos a todos aquellos modelos de metainformación que, lejos todavía de un nivel formal de normalización, han sido promulgados por una institución con cierta entidad en el ámbito informativo en que se crean; esto unido a su difusión en Internet, hace que se conviertan en estándares *de facto* reconocidos por el mercado *a posteriori*. Es el caso de modelos de metadatos nacionales que suponen la adaptación del DC o de otros formatos de propósito general, por ejemplo AGLS (*Australian Government Locator Service*⁴³) que, reconocido por las agencias gubernamentales y

⁴³ AGLS <http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html>. En el propio sitio Web de AGLS, donde se define esquema o conjunto de elementos de este formato, se refieren al [cont.]

desarrollado por el Archivo Nacional de Australia, es un estándar de metadatos a este nivel, implementado, como vimos, por algunas aplicaciones, como Metabrowser⁴⁴. Estos modelos implican una normalización tácita y consensuada entre un grupo de usuarios implicados que, en muchos casos, también asume el mercado, por lo que constituyen un tipo especial de estándares *de facto*.

Es preciso reflejar también que, a pesar de la normalización consorciada y de las tecnologías aparejadas a ella, el desarrollo de la WWW en los últimos años ha estado muy vinculado al desarrollo del mercado y a la implementación objetiva de las normas creadas de esta forma en las propias aplicaciones. En muchos casos, el desarrollo o más bien la utilización o aprehensión de los estándares para la Web, dependen de que las empresas creadoras de software para Internet (navegadores, buscadores y otras herramientas), implementen en el diseño de sus aplicaciones los estándares, por ejemplo del W3C⁴⁵. Se producen, así, situaciones de círculo vicioso o *chicken-and-egg*: los usuarios no utilizan una norma porque las aplicaciones no la interpretan, las aplicaciones no interpretan una norma porque los usuarios no la utilizan⁴⁶. El influjo del mercado en la regularización del uso de estándares es tan grande que, en ocasiones, incluso impone el uso de algún aspecto tecnológico que no

Dublin Core como un *estándar internacional de metadatos para la recuperación de información de recursos en línea (international online resource discovery metadata standard)*. Asumen pues, la internacionalización del DC aunque éste no haya sido todavía formalizado como estándar ISO. *Cfr. infr. 6.3.*

⁴⁴ *Cfr.* Capítulo 4, nota 184.

⁴⁵ Por ejemplo, en el caso del desarrollo de clientes de navegación, navegadores, visualizadores o agentes de usuario, no siempre se tienen en cuenta las normas emanadas del Consorcio Web, como ocurre *v. gr.*, con la interpretación de hojas de estilo (CSS2).

No obstante, existen ejemplos que indican que esta situación está cambiando. Por citar uno, el editor WYSIWYG que acompaña a la *suite* de programas del nuevo Netscape 6.x, Netscape Composer, no permite insertar una imagen en una página hasta que no se ha introducido un texto alternativo (atributo ALT del elemento IMG de HTML), lo que fuerza a que los autores de páginas Web hagan diseños más accesibles según las pautas del WAI (*Web Accessibility Initiative*) del Consorcio. *Vid.* World Wide Web Consortium. *Web Content Accessibility... Op. cit.*, <http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505>

⁴⁶ Hablaremos de este efecto al tratar el proceso de normalización del DC. *Vid. infr.* Nota 71.

ha sido normalizado *a priori*. Por ejemplo, la utilización de las metaetiquetas de HTML `<META NAME="keyword" CONTENT="...">` y `<META NAME="description" CONTENT="...">`, se popularizó porque algunos buscadores de Internet las reconocían para establecer sus *ranking* de relevancia, hasta tal punto que al hablar de modelos de metadatos algunos autores se referían a ellas como *etiquetas <META> de Altavista*, reconociéndolas al mismo nivel que otros estándares de metadatos en la Web⁴⁷.

Además de las instituciones encargadas de la normalización a nivel nacional e internacional, de consorcios de origen empresarial y de instituciones "ejemplares" de un ámbito informativo, el mercado en sí mismo se ha convertido en una máquina de proponer y de crear estándares y opciones tecnológicas. Así, no es extraño que distintas empresas individuales envíen notas, propuestas e incluso prototipos normativos al Consorcio Web⁴⁸. A pesar de todos estos mecanismos, más o menos reconocidos, para la popularización del uso de estándares, podemos distinguir, además, tres tipos de estándares⁴⁹ por su nivel de implantación. Dicho nivel de utilización debería tener correlación con su estatus normativo pero, en el ámbito de los metadatos y en la generalización de las tecnologías aplicadas al complejo mundo

⁴⁷ Tony Gill. Los metadatos y la World Wide Web. En: *Introducción a los Metadatos: vías... Op. cit.*, p. 17-18. Sin embargo, en la reedición de esta publicación en inglés en agosto del 2000, clasifica a este tipo de metadatos de propósito general basados en HTML como *Search Engine Metadata*, reconociendo que Altavista popularizó estas etiquetas. Tony Gill. Metadata and The World Wide Web. En: *Introduction to Metadata: pathways... Op. cit.*, http://www.getty.edu/research/institute/standards/intrometadata/2_articles/gill/index.html

⁴⁸ Recordemos por ejemplo, en el ámbito de los metadatos, iniciativas como el CDF o el MCF, propuestas directamente por Microsoft y Apple-Netscape, respectivamente. *Vid.* 5.2.1.1.

⁴⁹ Esta clasificación es la que utiliza la guía para la correcta utilización de la información electrónica publicada en 1997 por la UE para clasificar el nivel de implantación-utilización de cada uno de los estándares que se comentan en ella. Sin embargo, a pesar de que es una tipificación de didáctica implícita propia de un manual como el que citamos, nos parece muy oportuna para generalizar el nivel de normalización en la WWW así como para individualizarla al uso de estándares de metainformación vinculados a este entorno informativo. *Guidelines on Best Practices for Using Electronic Information... Op. cit.*, p. 10.

informativo de la WWW, depende más de la inercia, que del estatus real de un estándar o propuesta. De todas formas, se distinguen:

- a) Estándares reconocidos y estables. En teoría, el nivel de estabilidad de un estándar y su reconocimiento, debería estar vinculado directamente a la formalización, esto es, a la proclamación del mismo como una norma nacional o internacional. En el ámbito de Internet, en cambio, el proceso es inverso: primero se reconoce y se estabiliza su uso, y después se convierte en un estándar formal. Este proceso se pone de manifiesto, por ejemplo, en la trayectoria de desarrollo del DC. El índice de la estabilidad de una norma (formal, *de facto* o PAS) es su aprehensión por las herramientas de software, independientemente del estatus formal de la misma. Esta afirmación es evidente en el caso del HTML, cuya consolidación como estándar internacional coincide con su obsolescencia en la realidad⁵⁰.
- b) Estándares dependientes o utilizados sólo por un pequeño número de proveedores o empresas informáticas. Este nivel de implantación correspondería, en principio, a los "estándares formales de facto" o especificaciones de acceso público cuya implantación es potestativa para el mercado del software. La regularización creciente del uso de XML o la implantación generalizada, en años anteriores, de HTML en casi todas las aplicaciones, demuestran que el nivel de implementación de las tecnologías no está subyugado de su estado de normalización.
- c) Estándares propietarios cuyo uso no puede garantizarse. Este sería el caso de, por ejemplo, las HCL (listas de control de indización) que utilizan exclusivamente la herramienta para la edición y gestión de metadatos Metabrowser⁵¹, cuya generalización por el resto de aplicaciones relacionadas no puede garantizarse. Empero, efectos normativos de tecnologías propietarias como OpenURL de Ex-

⁵⁰ *Vid. supr.* Nota 19. *Cfr.* 5.1.2.

⁵¹ *Vid.* 4.3.2.2. y Glosario, Fig. 69.

Libris, comentado anteriormente, demuestran que la implantación de un modelo, formato o tecnología no depende tampoco de la filiación del mismo a un software particular.

Lo que marca la validez de una norma o lo que, incluso, la convierte en norma no es tanto su promulgación formal sino más bien su nivel de utilización y/o aplicación en el ámbito al cual va dirigida, en el caso que nos ocupa, el de los creadores de contenidos Web o el de los "bibliotecarios digitales". Además de su establecimiento como estándar universalmente aceptado y adoptado *de facto* o *de jure*, el seguimiento explícito del conjunto de reglas, usos y recomendaciones consignados en ella, tampoco depende de su formalización. Nuevamente, el ejemplo es HTML: es el lenguaje de marcado predominante de la actual WWW, de estructura rigurosa definida en SGML por el Consorcio Web Mundial como PAS y, desde el año 2000, estándar formal internacional de la ISO; sin embargo, el rigor en su utilización es muy laxo. Como dice Floyd⁵² *parece ser que todo el mundo respeta las normas como si las hiciera la Federación Mundial de Lucha Libre, es decir, nadie les hace caso.*

Frente a toda esta casuística, que relativiza el valor del establecimiento de estándares formales vinculados a la WWW, la tendencia es que las distintas tecnologías, los modelos de metadatos también, se legitimen a través de un proceso de normalización nacional e internacional. El efecto de acceso universal a la información que lleva aparejada la Web, hace que los desarrolladores de software y los gestores de información tengan que pensar "en internacional" para sobrevivir en el mercado global que se genera. Todavía hoy, la creación de productos y servicios que se funden en estándares internacionales garantizan, de alguna manera, la permanencia en el mercado informático; asimismo se promueven estándares formales que garanticen la pervivencia, en este caso, de un modelo de metainformación. Como

⁵² Michael Floyd. *Creación de sitios Web con XML*. Madrid: Pearson Education, 2000, p. 284.

reconoce François Horn⁵³ la elección de normalizar se puede interpretar como un *proceso de selección de un equilibrio de coordinación en el seno de múltiples equilibrios posibles*, pero esto no garantiza que el equilibrio que se consiga sea el más eficaz.

No todas las disciplinas que utilizan modelos de metadatos para la descripción de su información se pueden comprometer con largos procesos de estandarización formal debido a, entre otros condicionantes, la rápida evolución de Internet. Sí es necesario, de todas formas, que exista un organismo que centralice el acceso a la descripción e información relativa a cada sistema de metadatos y que mantenga, en cierta manera, la evolución del mismo. Esa institución será la que se encargue de mantener la normalización y el desarrollo del modelo durante ese estadio normativo que señalábamos antes, incipiente a las especificaciones de acceso público. De esta manera podemos destacar, verbigracia, OCLC que se ocupó de potenciar el DC hasta que se formó el grupo o consorcio de la *Dublin Core Metadata Initiative*, que se encarga, en la actualidad, de su desarrollo y normalización; la *Library of Congress* y la Asociación Americana de Archiveros, que se encargan de la misma forma de impulsar el EAD; o la Universidad de Virginia que realizaría el seguimiento de TEI hasta que se creó el consorcio para desarrollar esa DTD de SGML/XML. Incluso, a un nivel general, podríamos destacar el CERN que se ocupó del desarrollo de HTML antes de que se instituyera el W3C. Así, los sistemas de metadatos perpetrados por consenso se convierten fortuitamente en normas *de facto*, sobre todo, si se realiza el control y mantenimiento precisos, y si su desarrollo descansa en otros estándares⁵⁴.

⁵³ François Horn. Diversité des informations traitées par des moyens informatiques, standardisation optimale et acteurs su processus de standardisation [documento HTML]. *Solaris*, Decembre 1999/Janvier 2000. Disponible en: <http://www.info.unicaen.fr/bnum/jelec/Solaris/d06/6horn.html> (consultado el 25 de octubre de 2000).

⁵⁴ Por ejemplo Z39.50, que está estudiando la posibilidad de acomodar el sistema de metadatos del Dublin Core. Vid. Ralph Le Van. *Op. cit.*, <http://cypress.dev.oclc.org:12345/~rrl/docs/dublincoreandz3950.html> Asimismo, existe un grupo de trabajo para que Z39.50 interopere con metadatos RDF. Dan Brickley. *ANNOUCE: Mozilla RDF/Z39.50 Integration Project* [mensaje en [cont.]

Con el paso del tiempo, las distintas comunidades de interés o grupos de trabajo temáticos, que han desarrollado sus propios esquemas de descripción de metadatos, tratan de normalizar formalmente esos modelos en aras a facilitar el intercambio de información electrónica, inicialmente dentro de la misma área temática, que posteriormente, tenderá a traducirse en un intercambio de información interdisciplinar, gracias a la flexibilidad de XML, en la mayoría de los casos. La casuística del nivel y del estatus de normalización de metadatos es tan variada como la del resto de las normas surgidas a tenor de la Web y depende fundamentalmente, del tiempo que el formato lleve en desarrollo y del nivel de compromiso de los agentes o partes implicadas en su desarrollo (*stakeholders*⁵⁵).

El paso del tiempo de la creación de un modelo de metadatos, las experiencias de aplicación que generan masa crítica y la actividad de sus *stakeholders* en pos de la normalización, darán lugar a distintas formas de fijar el uso correcto o aconsejado de cada formato. Así, aparecen guías, directrices, reglas, recomendaciones, manuales de procedimiento (*best practice*⁵⁶), etc., e incluso normas propiamente dichas. Por ejemplo, el esquema (*schema*) EAD ha dado lugar a unas directrices de aplicación elaboradas por el *Encoded Archival Description Working Group* de la Sociedad Americana de Archiveros⁵⁷, mientras que el DC ha dado lugar a un estándar formal, como analizaremos en el apartado siguiente. En cualquier caso, independientemente del nombre puntual que se le dé a la definición de elementos que conforman un modelo y a las reglas para su aplicación, el carácter normativo de un esquema de

grupo de noticias]. Disponible en: netscape.public.mozilla.rdf, 24 de agosto de 1999; 01:04:50GMT. Cfr. <http://www.mozilla.org/rdf/doc/z3950.html> (consultado el 27 de agosto de 1999).

⁵⁵ Vid. Glosario.

⁵⁶ Vid. Glosario. Tanto a partir en la bibliografía de este trabajo, como de los URLs consignados en el índice de siglas y acrónimos recogido al final del mismo, se recogen diferentes ejemplos de esta casuística, así como de las otras formas de definir y divulgar esquemas o estándares de metadatos.

⁵⁷ Esta normativa se tradujo en España por la Fundación Tavera en el año 2000. Society of American Archivist. *EAD. Descripción Archivística Codificada: directrices de aplicación. Versión 1.0*. Fundación Histórica Tavera, ed., trad. [Alicante]: Fundación Histórica Tavera, 2000.

metadatos se hace sentir básicamente en dos aspectos: por un lado, en una utilización generalizada de la norma en cuestión, al menos en el ámbito informativo al que está dirigida; y por otro, en la creación de aplicaciones compatibles con dichas normas⁵⁸.

Toda la casuística de denominaciones, unida a la vocación normativa de todos ellos reflejada en la utilización y la creación de aplicaciones, hace que se intercambie con mucha versatilidad la denominación de estándar, modelo, formato, conjunto de elementos de metadatos, etc.⁵⁹. Salvo ante el hecho evidente de que se hayan convertido en un estándar formal internacional (ISO) o nacional (ANSI/NISO, en el caso americano) o en una norma regional (por ejemplo, en Europa, las normas del CEN), es difícil determinar el nivel exacto de normalización de un formato de metadatos.

SCHEMAS es el proyecto del programa IST de la Comisión Europea que trata de recoger esta realidad, constituyendo un forum para todas aquellas instituciones, compañías, etc. que están desarrollando iniciativas de metadatos en Europa, o que

⁵⁸ Por ejemplo, y siguiendo con el caso de las EAD, citado arriba, la institucionalización del modelo en alguna forma normativa (en este caso directrices) hace que las empresas productoras de software contemplen su implementación en los productos que desarrollan. Tal es el caso, por ejemplo, de Baratz, propietario de la herramienta de gestión integrada de documentación archivística Albalá, que ha anunciado la compatibilidad futura de su sistema con la norma EAD Bartz, para promocionar su producto de software, argumenta que su sistema soporta estándares internacionales para la descripción archivística —ISAD(G) e ISAAR(CPF)—, y que además, *...se han analizado las recomendaciones con relación a los procesos de descripción archivística elaboradas por el Consejo Internacional de Archivos (C.I.A.). En esta misma línea de adecuación e integración de recomendaciones y normas internacionales, se han estudiado también otros proyectos existentes para la codificación normalizada de los instrumentos de descripción --proyectos EAD--* Albalá: Sistema integrado de gestión de centros archivísticos [documento PDF]. Madrid: Baratz, Servicios de Teledocumentación, 2000. Disponible en: <http://www.baratz.es/baratz/pdf/albala.pdf>, p. 4 (consultado el 28 de octubre de 2000). Esto demuestra que la formalización de modelos de descripción en normas internacionales, implica una adopción inminente de ellos en las herramientas informáticas específicas y, recíprocamente, esa adaptación en el software de las normas fomenta la generalización de su utilización.

⁵⁹ Para no ahondar ahora en cuestiones terminológicas, realizamos una reflexión y justificación del uso de estos términos para designar los "estándares" de metadatos independientemente de su formalización, en el glosario explicativo de la terminología utilizada que incluimos en la parte final de este trabajo. Vid. Glosario, *Metadata [Format/Model/Registry/Schema/Standard]*.

pretenden desarrollar servicios de información en Internet de alta calidad basados en ellos. Dentro de SCHEMAS se da mucha importancia al marco de la normalización, procurando definir o ratificar distintas iniciativas de metadatos, como estándares. Se da cabida no sólo a los estándares formales, reconocidos así tradicionalmente por el hecho de haber sido promulgados en el marco de una institución normalizadora como la ISO, sino también a otras iniciativas basadas en la propia capacidad de Internet, que ha conferido, desde nuestro punto de vista, la potestad normalizadora a instituciones como el *Federal Geographic Data Committee* (FGDC) e inclusive a conjuntos de *stakeholders* formalmente reunidos, como en el caso de la DCMI. En el segundo informe sobre los estándares de metadatos, realizado por Dekkers y Eilert⁶⁰ en el seno de SCHEMAS, se categorizaban así, las veintidós iniciativas detectadas de normalización de metadatos con implantación en Europa:

- a) *Modelos*. Los modelos no son actividades de normalización propiamente dichas, sino que más bien describen modelos para la gestión de metadatos. Según estos autores, pertenecen a esta categoría: el CRM⁶¹ (*CIDOC object-oriented Reference Model*), <indec>, OAIS o el FRBR de la IFLA.

⁶⁰ Makx Dekkers, Christian Eilert. *Standards Framework Report 2* [documento PDF]. SCHEMAS, 27 de Mayo 2001. Disponible en: <http://www.schemas-forum.org/stds-framework/2.pdf> (consultado el 13 de octubre de 2001).

⁶¹ El CRM representa una especie de ontología para la información del patrimonio cultural. Describe en un lenguaje formal los conceptos implícitos y explícitos y sus relaciones en la documentación de patrimonio, tratando de proporcionar una base semántica para dar coherencia a ese tipo de información. Más información: http://cidoc.ics.forth.gr/docs/crm_version_3.0.rtf

El hecho de incluir este tipo de "metadatos" bajo la designación de "modelos", podría poner en tela de juicio el título que hemos dado al capítulo anterior (Modelos de metadatos), en el que analizamos algunos de los formatos de metainformación partiendo de la perspectiva del marcado descriptivo de la información. No obstante, la bibliografía avala esta denominación. Por ejemplo, Weibel se refiere así al DC. —*Vid.* Stuart Weibel. The Dublin Core: A Simple Content Description Model for Electronic Resources. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, vol. 24, nº 1, October/November 1997, p. 9-11—. Además, como ya hemos dicho, los metadatos son, ante todo, datos (*Cfr.* 3.1.). De forma genérica, desde un punto de vista informático, vinculado al mundo de las bases de datos, la denominación de "modelo de datos" se da a todo un sistema formal y abstracto que permite describir los datos de acuerdo con reglas predefinidas y que incluye la propia estructuración de los datos. Por ello, nos parece genérica y apropiada la

[cont.]

- b) *Estándares básicos*. En esta categoría reconocen actividades normalizadoras que proporcionan una infraestructura de trabajo para otras normas; así, pertenecen a ella RDF y la norma ISO 11179 de registros de metadatos⁶².
- c) *Estándares interdominios (cross-domain standards)*. Es la categoría que establece SCHEMAS *ex professo* para el Dublin Core, dada su capacidad de adecuarse a cualquier dominio informativo, a pesar de su condición original de ser un modelo de propósito general.
- d) *Estándares específicos de un dominio*⁶³. En este aspecto se reconocen estándares en seis dominios concretos:
- El dominio de la *publicación*, que cubre la distribución de música, vídeo, libros y multimedia, la industria de la música, revistas científicas, agencias de prensa, periódicos y gestión del copyright. Los estándares reconocidos dentro de esta categoría son ONIX (*Online Information eXchange*), EBX (*Electronic Book eXchange*) y DOI.
 - El dominio de lo que denominan en SCHEMAS, *audiovisual* y que abarca la industria fílmica, producción y archivo de reportajes y multimedia, etc. En este grupo se contemplan normas como MPEG-7 o SMPTE (*Society of Motion Picture and Television Engineers*).

denominación de "modelos" para referirnos a todas las iniciativas de metadatos que describen las reglas de estructuración de la información y los elementos apropiados para hacerlo. Por ello, también definíamos RDF como un "metamodelo", en el sentido de que proporciona un sistema único de integración de modelos de metadatos para estructurar la información, de forma semejante a los sistemas orientados a objetos.

⁶² ISO 11179: *Specification And Standardization of Data Elements*. Es la norma internacional básica diseñada para proporcionar un marco robusto en la definición de elementos de datos y también registros de metadatos. *Cfr.* Glosario, *Metadata [Format/Model/Registry/Schema/Standard]*.

⁶³ Esta diferenciación entre estándares de un dominio y estándares interdominios se corresponde básicamente a la que hicimos en el capítulo anterior para analizar los distintos modelos: metadatos de propósito general (5.2.1.) y metadatos de propósito específico (5.2.2.), pero se reconoce, con el calificativo "interdominios", la capacidad integradora que ha adquirido, sobre todo en los últimos dos años, el Dublin Core.

- *Patrimonio cultural*. En esta categoría se incluyen fundamentalmente, los estándares de bibliotecas, archivos y museos (como MARC21, EAD o SPECTRUM).
- *Educación*. Se incluyen en este grupo iniciativas relacionadas con conceptos difundidos en los programas europeos de la Sociedad de la Información del ámbito educativo, como la tele-educación o el aprendizaje a lo largo de toda la vida. Así, pertenecen a esta categoría normas como el IMS, IEEE y LOM.
- *Dominio académico e investigación*, que implica los estándares desarrollados en o para contextos de investigación, sociedades profesionales etc., como CERIF (*Common European Research Information Format*) o la propia Iniciativa de Codificación Textual, TEI.
- La última gran categoría recoge el ámbito *gubernamental, geoespacial y medio ambiente* y, de forma genérica, todos los estándares relacionados con el sector público, por ejemplo: GILS, FGDC, o GELOS.

Todo estándar, también los de metadatos, tiene un ciclo de vida o evolución que generalmente comprende, al menos siete fases que comentamos a continuación. Para referirnos al ciclo de vida de los estándares de metadatos tomamos como modelo de referencia RDF, ya que es un estándar básico en la clasificación de SCHEMAS, y el valor prospectivo de dicha especificación emanada del Consorcio Web (PAS) es, a nuestro juicio, como anticipábamos en el capítulo anterior, ser "el" estándar para la codificación y gestión de metainformación. Siguiendo a la WSSN⁶⁴, las distintas etapas de una norma son:

⁶⁴ World Standards Service Network. *Op. cit.*, http://www.wssn.net/WSSN/gen_inf.htm#Standards

- 1) *Identificación de las necesidades* y de la viabilidad técnica y económica del trabajo normativo. En esta fase se debe responder a dos cuestiones: *¿la norma proporcionará un "plus" económico y técnico al sector? ¿Existe el conocimiento necesario requerido para llevar a cabo la norma?*

Esta pregunta, desde nuestra interpretación, llevada al mundo de los metadatos, está íntimamente relacionada con la existencia o no de masa crítica y con la base de conocimientos sobre el modelo de metadatos que puedan aportar los *stakeholders* del sector. Así por ejemplo, en el caso de lo que luego será RDF, el grupo de agentes influyentes en el ámbito de los metadatos había detectado ya en 1996 el valor técnico de contar con un marco contenedor para los metadatos (*Warwick Framework*).

- 2) *Programación colectiva*. En esta fase es necesario determinar las bases de las necesidades y prioridades, para posteriormente registrar un programa de trabajo en la organización implicada.

Según la idea de SCHEMAS, y en el mismo sentido de todos los comentarios que hemos apuntado en torno a la cuestión de los estándares de metadatos, la institución implicada en el proceso normativo no tiene por qué ser una institución formal de normalización (ISO, ANSI); puede ser también un consorcio (como el W3C en el caso de RDF, o el TEI-C en el caso de la TEIH).

- 3) *Las partes interesadas redactan de un borrador de la norma*. Normalmente ésta es una actividad delegada a un grupo de personas, que en el caso del W3C serán los editores de la norma. Por ejemplo Ora Lassila y Ralph Swick serán los editores del modelo y la sintaxis de RDF⁶⁵, sin embargo, desde el primer borrador

⁶⁵ RDFMS, *Op. cit.*, <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>

(1/08/97), se recoge la opinión de todo un grupo de trabajo que representa distintos grupos interesados vinculados a RDF⁶⁶.

- 4) *Discusión y consenso de los expertos sobre el borrador.* Sin que ello obste para que se celebren congresos, conferencias, talleres, etc., la discusión en el ámbito de Internet se realiza básicamente a través de las listas de distribución (en el caso de RDF, por ejemplo: *www-rdf-comments*, *www-rdf-interest*), *newsgroup* e incluso, cada vez más en el W3C, discusiones en tiempo real a través de IRC.
- 5) La etapa de *validación*, que implica la consulta a nivel internacional o nacional gracias a las convocatorias públicas. Esta etapa supone, en el caso del W3C la reelaboración y publicación de distintos borradores. Por ejemplo, en el caso de RDF se hicieron diecisiete revisiones formales desde la publicación del primer borrador público (2/10/97) hasta el último borrador, para proponerlo como recomendación el 19 de diciembre de 1998.
- 6) *Aprobación del texto para la publicación como estándar.* En el Consorcio Web, esta fase conlleva tres tipos de documentos: recomendación candidata (CR), propuesta de recomendación (PR) y, finalmente, recomendación (REC).
- 7) De forma semejante a lo que ocurre con los procesos legislativos, la elaboración de las normas implica una "regulación" posterior y una revisión de las mismas. En el caso de RDF, esa revisión constante se manifiesta, por ejemplo, en la

⁶⁶ Los miembros del Grupo de Trabajo de la especificación del modelo y la sintaxis de RDF (RDFMS, *Ibid.*) que ayudaron a diseñarla, debatir propuestas, corregir numerosos borradores y finalmente, conseguir consenso, son 27 personas de distintas instituciones con intereses en los metadatos, como empresas de software, servicios de información o del ámbito de la investigación universitaria. *V. gr.* Ron Daniel (Datafusion), Renato Iannella (DSTC), Chris McConnell (Microsoft), R.V. Guha (Netscape), Ora Lassila (Nokia), Eric Miller (OCLC), Charles Tim Bray (Textuality), Tim Berners-Lee (W3C), Dan Brickley (UK Bristol), etc. Todos ellos pertenecen al grupo de *trabajo RDF Core Working Group* <<http://www.w3.org/2001/sw/RDFCore>>.

corrección de erratas⁶⁷ o en la formulación de propuestas posteriores relacionadas con la especificación⁶⁸.

El rigor de todas estas etapas (prácticamente idénticas a las llevadas a cabo para el establecimiento de una norma formal en sentido estricto), junto al carácter internacional del W3C, es lo que nos hace hablar de las especificaciones de acceso público que emanan de él como normas "*formales de facto*", que se han convertido en una forma ágil y escrupulosa de regular la WWW.

Al igual que el W3C, otros consorcios e instituciones han arbitrado sus propios métodos para establecer el *iter* normativo de sus especificaciones, reflejando en todo momento el estado gradual de su constitución. Así, por ejemplo, el IETF (*Internet Engineering Task Force*) tiene establecidos distintos tipos de estadios de sus normas: *Internet Draft*, *Request for Comment*, *Proposed Standard* y finalmente *IETF Standard*, que constituyen en muchos casos un punto de referencia para la

⁶⁷ Ralph R. Swick. *Errata in REC-rdf-syntax-19990222* [documento HTML]. W3C, 31 de mayo de 2001. Disponible en: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/errata> (consultado el 8 de agosto de 2001).

⁶⁸ En el caso de RDF, estas revisiones constantes irán desde el *Cambridge Communiqué*, público a los siete meses de la propia especificación que alentaba el compromiso del trabajo sobre las relaciones de los esquemas RDF y XML, a las nuevas propuestas, como la gramática para la sintaxis RDF/XML, la teoría del modelo RDF o el test de casos RDF. *Vid.*

- World Wide Web Consortium. *The Cambridge Communiqué* [documento HTML]. Ralph R. Swick, Henry S. Thompson, eds. W3C, 20 de septiembre de 1999. Disponible en: <http://www.w3.org/1999/08/27-communicue-19990920> (consultado el 30 de septiembre de 1999).
- World Wide Web Consortium. *Refactoring RDF/XML Syntax. W3C Working Draft 06 September 2001* [documento XHTML]. Dave Beckett, ed. W3C, 6 de septiembre de 2001. Disponible en: <http://www.w3.org/TR/2001/WD-rdf-syntax-grammar-20010906> (consultado el 7 de octubre de 2001).
- World Wide Web Consortium. *RDF Model Theory. W3C Working Draft 25 September 2001* [documento XHTML]. Patrick Hayes, ed. W3C, 25 de septiembre de 2001. Disponible en: <http://www.w3.org/TR/2001/WD-rdf-mt-20010925> (consultado el 7 de octubre de 2001).
- World Wide Web Consortium. *RDF Test Cases. W3C Working Draft 15 November 2001* [documento XHTML] Art Barstow, Dave Beckett, eds. W3C, 15 de noviembre de 2001. Disponible en: <http://www.w3.org/TR/2001/WD-rdf-testcases-20011115> (consultado el 21 de noviembre de 2001).

estandarización posterior. Otras instituciones más destinadas al mercado y por tanto, a la formulación de estándares *de facto* en sentido estricto, también han adoptado modelos semejantes de evolución. Tal es el caso por ejemplo de OASIS (*Organization for the Advancement of Structured Information Standards*), un consorcio internacional sin ánimo de lucro que diseña y desarrolla estándares industriales para la interoperabilidad basada en XML. Sus normas presentan dos estadios: TR (*OASIS Technical Report*) y *OASIS Standard*. Este consorcio está formado por importantes empresas del sector como Sun, HP, Informix, Oracle, Documentum, SoftQuad, Microsoft, Enigma, Open Network Technologies, Fujitsu, Arbortext, etc., y ejerce un gran influjo en el desarrollo de estándares para la WWW, sobre todo, desde el punto de vista de la actualidad imperturbable de la información que brinda.

Todas las reflexiones que hemos apuntado, unidas a que el observador y promotor europeo de los estándares de metadatos (SCHEMAS) reconoce como tales cualquier iniciativa con cierta pretensión de estabilizar un modelo de metainformación, podemos decir que la casuística de la normalización de metadatos es diversa y, en ocasiones, desde el punto de vista de su utilización, paradójica. No obstante, desciframos de todo ello que, para que los metadatos consigan un grado aceptable de eficacia y aplicación, es necesario que cuenten con lo que denominábamos en el apartado anterior "entornos normalizados", ya sean estándares formales, especificaciones de acceso público o estándares *de facto*. Empero, la tendencia latente marcada por procesos de normalización formal como el del FGDC o el DC, es que se conviertan en estándares internacionales, nacionales o regionales (ISO, ANSI, o CEN). Mientras tanto, instituciones y consorcios, como el W3C, marcan la pauta de los modelos de información sobre la información electrónica y por tanto aplicables también a bibliotecas digitales cuyos fondos son exclusivamente DLOs.

6.3. Desarrollo de un estándar formal: el caso del DCMES

A pesar de que los consorcios y las distintas instituciones tienen sus propios métodos para la formulación de estándares, la mejor forma de examinar, en la realidad de su uso, la carga normativa de un formato de metadatos consiste en analizar su nivel de utilización o conocer cómo ha sido su desarrollo en el tiempo. Para ello, es especialmente significativo tratar el caso paradigmático del conjunto de elementos del Dublin Core (DCMES) que empezó siendo un formato más para la organización de la información electrónica, surgido a mediados de los 90's, y que se ha convertido recientemente en un estándar americano ANSI/NISO. A través de este proceso del DC se podría justificar, pese a todas las reflexiones y opiniones que hemos apuntado, la creación de estándares formales (internacionales o nacionales) para los metadatos.

La necesidad de contar con normas formales para el formato Dublin Core se manifestó expresamente en el DC-5 celebrado en Helsinki en 1997. Entonces era demasiado pronto para comprometerse en la complejidad de la formalización del estándar y se dudaba mucho de su estabilidad. En ese mismo año, durante el primer taller de metadatos celebrado en Luxemburgo dentro del programa europeo *Telemática para Bibliotecas* de la DGXIII, se advirtieron varias carencias del DC que ponían en entredicho la necesidad y/o la oportunidad de una norma formal. De entre las críticas que recibió dicho formato de metainformación, podemos destacar algunas como la falta de estabilidad, la falta de información definitiva, de despliegue, la ausencia de herramientas de creación e indización basadas en este formato, la carencia de protocolos que soportaran dicho modelo, así como de formatos de intercambio, la falta de registros y de directrices claras de desarrollo, e incluso, la falta de representación de varios sectores que deberían estar implicados o, al menos, interesados en el Dublin Core. Las conclusiones del primer taller europeo sobre

metadatos argumentaron todas estas carencias del formato DC en los siguientes términos⁶⁹:

- No existía (en diciembre de 1997, cuando se celebró el encuentro de Luxemburgo) una responsabilidad formal para el mantenimiento del DC. Los desarrollos, hasta entonces, sólo tenían lugar en un grupo informal de expertos que se encontraban una o dos veces por año en lo que se conoce como la serie de encuentros del Dublin Core (*Dublin Core Workshops*).
- El estado técnico era inestable. Se denunciaba que durante los encuentros del grupo del Dublin Core se hacían cambios al formato y no existía una convergencia clara hacia una versión estable.

Recordemos por ejemplo, que en el DC-4, celebrado en marzo de aquel mismo año, surgió toda la contienda entre minimalistas y estructuralistas dando lugar a dos versiones del Dublin Core, *qualified* y *unqualified*. Aunque pese a esa discusión posteriormente se demostraría la versatilidad del DC, por aquel entonces, denotaba una falta de consenso.

- Se puso de manifiesto que el uso del formato de metadatos del DC no estaba apoyado en la existencia de recomendaciones y que su filosofía y su terminología no eran obvias para los usuarios no iniciados. Lo que podía llevar a interpretaciones diversas o divergentes que afectarían negativamente a la interoperabilidad.
- Se percibió que la aceptación del DC era lenta y también se advirtió la ausencia de masa crítica. Esta opinión surgida del encuentro de Luxemburgo sobre metadatos era suscrita, además, por otros autores a título individual. Tony Gill⁷⁰,

⁶⁹ *Metadata Workshop: Report of the Workshop Held in Luxembourg, 1 and 2 December, 1997*. [s.l.]: European Commission, Directorate General XIII – E/4, February 1998, p. 9-10, 13.

⁷⁰ Tony Gill. *The Dublin Core Metadata Element Set... Op. cit.*, p. 46.

por ejemplo, en un irónico pero interesante artículo publicado en el mismo año 97, daba una visión escéptica de este formato identificándolo con una situación clásica *chicken-and-egg*⁷¹.

Pese a todas estas críticas, ya se reconocía entonces la buena posición del DC para constituir un denominador común de interoperabilidad entre descripciones de recursos de la Red a través de todos los sistemas, disciplinas y culturas; permitiendo a los profesionales de la información proporcionar un acceso mediatizado al vasto depósito documental de información útil existente, que calificaba también en sentido positivo, la situación del DC.

Por otra parte, aunque también con sentido crítico, el sector bibliotecario más tradicional temía que la formalización del DC pusiese en peligro la catalogación como actividad fundamental bibliotecaria. Así, por ejemplo, Atting⁷² denunciaba la ausencia de reglas y principios rectores que aseguraran la calidad del contenido de un conjunto de metadatos asignado según el Dublin Core, y consideraba que este formato estaba dirigido a un conjunto más amplio y menos exigente de productores de información que las Reglas de Catalogación. Una descripción de un recurso basada en el DC sería, para este autor, *más incompleta y menos rigurosa en su*

⁷¹ *Cfr. supr.* Esta definición, tan angloamericana, de la situación del Dublin Core expresa lo que en español coloquial se diría "la pescadilla que se muerde la cola" para referirse a situaciones irresolutas por entrar en una inercia cíclica de las partes implicadas (en este caso los buscadores y los creadores de recursos). La calificación de la situación del Dublin Core de *chicken-and-eggs* (qué fue primero el huevo o la gallina) se basa en la observación de que los autores y editores no invertirán tiempo en crear metadatos basados en el DC si los software/robots que recorren e indizan la Web no los utilizan. Y los motores de búsqueda genéricos de Internet, a su vez, no realizarán una indización selectiva basada en este formato de metadatos, si los creadores de información Web no asignan metadatos-DC, de forma suficientemente extendida para que se indice en virtud de ellos.

⁷² Jonh Atting. Dublin Core Metadata and the Cataloging Rules [documento HTML]. En: American Library Association. *Committee on Cataloging: Description and Access Task Force on Metadata and Cataloging Rules: Final Report*. ALA, 3 de junio de 1998, rev. 4 de junio de 1998. Disponible en: <http://www.ala.org/alcts/organization/ccs/ccda/tf-tei5.htm> (consultado el 18 de junio de 1998). Estas apreciaciones de Atting darían lugar a una amplia discusión en torno al nivel de normalización del contenido de los datos de un registro basado en el formato DC, aspecto que trataremos parcialmente en el apartado siguiente (6.4.).

contenido que un registro tradicional MARC, y sólo serviría como fuente de catalogación a la hora de incluir el registro de un recurso Web en un catálogo. Esta postura de Attig respecto al DC responde, en nuestra opinión, al desvelo catalográfico que ha infundido siempre al trabajo bibliotecario, al recelo de asimilar nuevos estándares y a la resistencia al cambio.

Todas estas críticas no impidieron que, a partir de 1998, se iniciara un foro de debate abierto en el seno de la DCMI⁷³, que demostraba la impaciencia en la normalización formal del *Unqualified Dublin Core*. En septiembre de 1998 el conjunto de elementos del DC ya era un RFC del IETF⁷⁴ y a finales de ese año (4/11/1998), justo antes de la celebración del sexto congreso del Dublin Core, se constituyó oficialmente un grupo de normalización con el siguiente objetivo:

Ser un forum para mantener en contacto a las personas implicadas en distintos esfuerzos de normalización y seguir el progreso del Dublin Core, al mismo tiempo que se construye el camino de diferentes procesos de normalización. Este grupo está específicamente preocupado por los temas relacionados con la formalización del Dublin Core, y su aceptación dentro de iniciativas normalizadoras como las del CEN, ISO y NISO⁷⁵.

Ese mismo grupo emitiría un informe⁷⁶ dentro del propio DC-6 celebrado en Washington pocos días después de su constitución (8/11/98), justificando la formalización del DC e incidiendo en los aspectos de dicho formato criticados el año

⁷³ Cfr. Todos los mensajes de enero de 1998, bajo el subject: *RFCs and NISO Standardization of the DC*, en la lista de distribución: <http://www.mailbase.ac.uk/lists/dc-general>

⁷⁴ Stuart Weibel, et al. *Dublin Core Metadata for Resource Discovery...* Op. cit., <http://www.ietf.org/rfc/rfc2413.txt>

⁷⁵ DCMI Standards Working Group: <http://www.dublincore.org/groups/standards>

⁷⁶ Priscila Caplan, et al. Report on the Break-out Group on Standardisation [documento RTF]. En: *DC-6 Workshop (1998. Washington)*, 3 de noviembre de 1998. Disponible en: <http://www.purl.org/DC/workshops/dc6conference/pp/ws-dc6-breakout-standardisation.doc> (consultado el 3 de septiembre de 1999).

anterior. Consideraban que una norma formal serviría básicamente para crear una versión estable, ya que muchas instituciones estaban esperando esa estabilización para invertir en la creación de servicios de información basados en el DC. Además, un estándar formal (nacional —ANSI— o internacional —ISO—) justificaría su implementación en aquellas organizaciones que, como las agencias gubernamentales americanas, necesitan utilizar normas (internacionales) para sus actividades. Por todo ello, se decidió formalizar la versión 1.0 del DC, cuyo consenso ya había sido estipulado a través del RFC 2413. Para ello se enviaría la propuesta de normalización a tres instituciones con capacidad de establecer estándares formales: una a nivel internacional (ISO), otra a nivel regional (CEN) y otra a nivel nacional norteamericano (ANSI)⁷⁷.

Aparte de los aspectos criticables o desventajas del DC que ya hemos señalado con relación a los procesos de formalización de estándares de metadatos, el grupo de normalización del DC detectó otros problemas concretos, planteando empero, soluciones similares a las que se habían tomado previamente en el proceso de normalización formal del protocolo Z39.50. Siguiendo a Caplan⁷⁸, los problemas advertidos y las soluciones propuestas eran fundamentalmente, tres:

- 1) Las políticas de copyright de las instituciones responsables sobre los estándares que publican. La solución consistía en compartir los derechos para que la institución normalizadora pudiera vender el estándar al mismo tiempo que la DCMi se reservaba el derecho de difundirlo libremente en la Web.

⁷⁷ Concretamente, la propuesta de normalización del conjunto de elementos del DC 1.0 (DCMES) se envió al ISO TC46/SC4/WG7 (es decir al Comité Técnico 46 de la ISO encargado de las aplicaciones informáticas en información y documentación), al taller sobre metadatos para la información multimedia (MMI) del CEN/ISSS y al comité de desarrollo de normas NISO de la ANSI.

⁷⁸ Priscila Caplan, et al. Report on... *Op. cit.*, <http://www.purl.org/DC/workshops/dc6conference/pp/ws-dc6-breakout-standardisation.doc>

- 2) Las reglas rigurosas que tienen las instituciones normalizadoras en cuanto al formato de los estándares que publican. Para ello decidieron demandar una flexibilidad semejante a la tomada en el Z39.50, donde el contenido de las normas NISO e ISO es idéntico.
- 3) El mantenimiento de la norma formal. En este tema, se propuso que la agencia responsable del mantenimiento fuese la DCMI a través de su oficina de dirección, de sus comités asesores y de los propios *workshops*.

Con todo, la normalización ANSI/NISO prosperó y en enero de 2001 se hizo público el borrador final del texto; en octubre de ese mismo año, tras un periodo de consideración (febrero-marzo⁷⁹) y un periodo de votación (julio-agosto), se publicó el estándar Z39.85⁸⁰, donde se establecen de manera formal los quince elementos del Dublin Core, lo que puede ser *un primer paso para la normalización internacional*⁸¹.

Este proceso de normalización nacional del DCMES evidencia la lentitud de la formalización de metadatos. Cinco años han hecho falta para que sea una norma ANSI/NISO mientras que, como señalábamos antes, RDF llegó al grado de una Recomendación del W3C en poco más de dos años y su validez y aplicabilidad *de facto* es perfectamente equiparable, como demuestra ya la práctica y evolución de la WWW, a la de una norma formal.

⁷⁹ Los dos periodos de votación fueron del 2 de febrero al 5 de marzo y del 1 de julio al 15 de agosto. *ANSI/NISO Z39.85-200x: The Dublin Core Metadata Element Set. Status: Reconsideration ballot February 2-March 5, 2001* [documento PDF]. Bethesda, Maryland: NISO, rev. 1 de febrero de 2001. Disponible en: <http://www.niso.org/pdfs/Z3985-REV.pdf> (consultado el 18 de febrero de 2001).

⁸⁰ *ANSI/NISO Z39.85-2001. Op. cit.*, <http://www.niso.org/standards/resources/Z39-85.pdf>

⁸¹ Rebecca S. Guenther. Re: RFCs and NISO Standardization of the DC. [correo electrónico en lista de distribución]. *META2*, 13 de enero de 1998; 13:56. Disponible en: <http://www.roads.lut.ac.uk/lists/meta2/1998/01/0013.html> (consultado el 4 de agosto de 1998); <http://www.mailbase.ac.uk/lists/dc-general/1998-01/0013.html> (consultado el 26 de noviembre de 2001).

La formulación del DCMES como estándar formal de metadatos ANSI/NISO Z39.85-2001, es más una consecuencia que una causa del éxito y aplicabilidad del Dublin Core. El hecho de que el DC sea el formato de referencia para muchos otros modelos (como EdNA o AGLS), que las aplicaciones informáticas lo contemplen de forma creciente (como demostramos en el apartado 4.3.), que el Consorcio Web se refiera a él como semántica por defecto para el modelo y la sintaxis de RDF y en definitiva, la supremacía y utilización de la DCMI para la creación de servicios de información de calidad, bibliotecas digitales y/o *subject gateways*, se debe, desde nuestro punto de vista, a una serie de características y circunstancias que han ido confluyendo en este formato desde sus orígenes en 1995, que resumimos a continuación:

- Se trata de una iniciativa desarrollada por un grupo multidisciplinar de creadores de servicios de información que han sabido significarse ante la comunidad internacional de la información electrónica.
- La aplicación estratégica más importante de este conjunto de elementos es la capacidad de coherencia en la evolución. En un principio, este formato se dirigía a la descripción de recursos Web mediante un conjunto de elementos embebidos en la propia sintaxis del lenguaje HTML; a partir del año 1999, con la REC-W3C de RDF, ha adaptado su sintaxis a XML. Calificamos esta característica de estratégica porque, a pesar de que los metadatos insertados en el HTML no han sido nunca el único objetivo de la descripción basada en el DC, ha sido algo esencial para soportar su propio mecanismo y promover un despliegue rápido del conjunto de elementos que conforman este modelo, que conducirá a una utilización más profunda y consistente a tenor de RDF.
- Los elementos que constituyen el DC son fáciles de entender y sencillos de usar. Por ello, cualquier creador/autor de un recurso de la Red está en condiciones de describir su propio trabajo sin una formación específica previa. Este primer principio que imbuje el desarrollo del DC como formato de descripción de recursos en Internet, se incentiva, como vimos anteriormente (4.3), por la

proliferación de aplicaciones informáticas sencillas para la asignación de metadatos según este esquema; a su vez, las aplicaciones de metadatos contemplan cada vez más este formato por su uso creciente.

- El conjunto de elementos del DC tienen una vocación multidisciplinar (o interdominios, como calificaba SCHEMAS) e internacional en cuanto a su alcance y aplicabilidad. Por ello, decimos que se trata de un formato de metadatos de propósito general frente a otros modelos que están dirigidos a un tipo de información o materia específica..
- El DC es extensible, en el sentido de que permite agregar modificaciones aplicadas a una disciplina o trabajo específico, e incluso el propio DC está sujeto a revisiones y cambios, como demuestra su propia trayectoria. Permite que diferentes áreas de información traten de adaptar este formato a sus necesidades⁸².
- Todos los elementos son opcionales y repetibles. Esto supone una gran flexibilidad en la aplicación del formato y permite establecer distintos niveles de detalle en la descripción. Esta circunstancia, como hemos mencionado y como demostraremos al enfrentar un proyecto de aplicación real⁸³, constituye uno de los problemas esenciales a la hora de establecer colecciones digitales heterogéneas.

⁸² Sobre este aspecto en particular, Dempsey y Heery advirtieron, al principio, cierta tensión por extender los elementos del DC de tal forma que hicieran posible descripciones más complejas para dominios o disciplinas concretas, o para que su uso se extendiera a otro tipo de información distinta de los recursos Web, como por ejemplo al material impreso. Lorcan Dempsey and Rachel Heery. *Metadata: a current view... Op. cit.*, p. 163. Esta suspicacia de Dempsey y Heery en 1998 no se ha demostrado en la evolución natural de este formato ya que su perfecta integración con RDF, así como la posibilidad de crear perfiles (*application profiles*) lo han hecho adaptable a cualquier tipo de aplicación sin perder su fundamento como modelo de propósito general.

⁸³ *Cfr.* Capítulo 9.

- Por otra parte, y desde un punto de vista bibliotecario, el Dublin Core soporta las cuatro operaciones que realiza un usuario (o que espera poder realizar) ante un sistema de recuperación de información: búsqueda, identificación, selección y recuperación. De esta afirmación se deduce que el DC es un formato inspirado en la práctica bibliotecaria tradicional, y esta tendencia se manifestará también en todos sus desarrollos.
- Finalmente, el Dublin Core, partiendo de su cobertura geográfica internacional y su alcance interdisciplinar, proporciona una base para la interoperabilidad semántica entre dominios de metadatos más ricos. En otras palabras, los registros basados en formatos más ricos pueden *mapear*⁸⁴ el núcleo de los datos más importantes al Dublin Core, en aras a facilitar un conjunto de datos común para todas las expectativas de recuperación de información en Internet, sobre todo en el ámbito de las bibliotecas digitales e híbridas.

Por último, la DCMI, además de la formalización del conjunto de elementos como estándar ANSI/NISO Z39.85, tiene sus propios canales de normalización interna a través de especificaciones de dominio público o estándares formales *de facto*⁸⁵. Esta documentación normativa también muestra el estado evolutivo y el nivel de estabilidad de las especificaciones, gracias a varios tipos de documentos semejantes a los que establece en W3C (notas, borradores de trabajo, propuesta de recomendación y recomendación). La existencia de este desarrollo normativo paralelo a la estandarización formal demuestra, por un lado, la institucionalización del modelo de consorcio para el desarrollo de esquemas de metadatos, y por otro, la insuficiencia, debida a la lentitud, de la normalización formal.

⁸⁴ Vid. Glosario, *mapping*.

⁸⁵ Para ver la evolución normativa interna de la DCMI, Vid. <http://www.dublincore.org/documents>

6.4. Aproximación a la normalización del vocabulario

Uno de los problemas principales y tradicionales en la construcción y uso de sistemas de organización y recuperación de información ha sido la comprensión de los documentos y la representación de sus materias. A pesar de que los avances tecnológicos pueden proporcionar métodos para un acceso más rápido, el problema de la representación de la información sigue debiéndose a cómo se entienden e interpretan los documentos⁸⁶. Por otra parte, el problema intelectual de la caracterización del contenido ha constituido siempre una rémora en el trabajo documental; no obstante, asegurar la consistencia y el control del vocabulario han sido temas de gran preocupación para los bibliotecarios y tienen papel igualmente importante en el contexto de los metadatos para contribuir a la coherencia general de los resultados en la recuperación.

Entender la naturaleza del proceso de indización es más crítico ahora que nunca ya que el acceso temático a los documentos electrónicos puede proporcionarse tanto por una indización manual como a través de la indización automatizada⁸⁷, amén de que la información debe de ser igualmente legible por máquina y por el ser humano. Con esto no queremos decir que se trate de convertir el proceso de asignación de metadatos en un proceso técnico tradicional, donde el análisis de contenido sea una labor de análisis semiótico y una caracterización a través de un vocabulario controlado, pero debemos, al menos, mencionar las principales directrices relacionadas con el concepto de normalización del contenido y/o del vocabulario para sistemas de recuperación de información en red. De hecho, los detractores de los metadatos como fundamento para establecer sistemas organizados de información o

⁸⁶ Como ya explicamos, a la hora de utilizar una herramienta de extracción y asignación automática de metadatos como Klarity, el mayor problema en la automatización de la asignación de metainformación era justamente el contenido de las materias (Cfr.4.2.).

⁸⁷ En el capítulo siguiente trataremos algunas de las soluciones que han mediado los Sistemas de Recuperación de Información de Internet (SRII) a tenor de ambas formas de indización, su casuística y su relación con la metainformación.

bibliotecas digitales basan sus críticas en la ausencia de normalización del contenido y no en la coherencia de la estructura de los metadatos. Así por ejemplo, Michael Gorman opina que *los metadatos son un subconjunto del registro MARC sin las instrucciones sobre normalización del contenido, necesarias para crear un sistema de control bibliográfico*⁸⁸.

Por ello, en este apartado trataremos de esbozar otro aspecto de la normalización como son los estándares de valor semántico (tesauros, ontologías, clasificaciones, etc.), así como una somera aproximación a su utilidad en entornos de información a texto completo, cuya descripción responde a una estructura de metadatos. Conscientes de la complejidad que entraña analizar el tema de la recuperación por materias, para el que existe una bibliografía abundante, nuestra intención es sólo plantear genéricamente el problema.

Los sistemas de metadatos difieren en dos aspectos principales: la estructura y el contenido. El problema de la normalización de la estructura lo hemos tratado en los apartados anteriores a éste al juzgar su nivel de estandarización técnica como vocabularios *de elementos* que constituyen un modelo estructural y semántico de metadatos. Ahora nos ocuparemos del problema del contenido, que afectará a la interoperabilidad entre sistemas de recuperación de información y bibliotecas digitales desde el punto de vista de la necesidad del control de la designación, y del problema de establecer *normas sobre el valor de los datos*⁸⁹. Estas normas del valor de los datos controlan los términos o las palabras que deben completar una estructura

⁸⁸ Michael Gorman. From Card Catalogues *Op. cit.*, http://www.unicamp.br/bc/gorman_paper.html
Este tipo de críticas a los metadatos, provienen siempre de los sectores bibliotecarios tradicionales. Así por ejemplo, el Informe final del Comité de Catalogación de la ALA en torno a los metadatos matizó la inadecuación de éstos para el uso directo en los catálogos de bibliotecas porque: *no usan estándares semánticos o de contenido ni registros de datos como ficheros de autoridades, ni vocabularios controlados de materias*. American Library Association. *Committee on Cataloging: Description and Access*. *Op. cit.*, <http://www.ala.org/alcts/organization/ccs/ccda/ta-tei7.htm>

⁸⁹ *Introduction to Vocabularies: Enhancing Access to Cultural Heritage Information*. Los Angeles: J. Paul Getty Trust, 1998, p. 19.

de datos o metadatos. Pertenecen a este tipo de estándares las listas de autoridad y los vocabularios *de conceptos*⁹⁰. Debemos mencionar también en este sentido, los metadatos basados en materia, esto es, aquellos datos que representan las materias y sus interrelaciones que normalmente designan recursos de información específicos que pertenecen a esas materias, a esta clase de metadatos pertenecen las ontologías⁹¹ y los mapas de materia y su expresión en XML permite modelar las materias como propiedades de los recursos de información asociados a ellas.

Partimos de que la eficacia del sistema de metadatos para la recuperación de información electrónica por materias no dependerá sólo de la estructura coherente del modelo, sino de los valores de contenido que se añadan a esa estructura, esto es,

⁹⁰ Nótese la diferenciación que hacemos entre las normas destinadas a establecer formatos de metadatos, esto es, vocabularios *de estructuras semánticas*, frente a los vocabularios *de conceptos* (tesauros, glosarios, listas de materias, etc.).

⁹¹ Las ontologías son un nuevo concepto que ha surgido, independientemente de su significado filosófico tradicional, vinculado a varias iniciativas relacionadas con la Web semántica. Aunque ya la hemos definido (*Cfr.* definición de Gruber, capítulo 3, nota 101) una ontología es, en términos muy generales, un sistema semántico que contiene tanto los términos, como sus definiciones y una especificación de las relaciones entre dichos términos. Las ontologías son, desde nuestro punto de vista, una visión moderna, informatizada y más potente, de la utilización de los vocabularios temáticos controlados para la recuperación de información en entornos de red y que se traducirá en una mayor precisión semántica.

Uno de los proyectos más notables de ontología fue CYC® <<http://www.cyc.com>>, desde entonces, las posibilidades de los lenguajes de marcado y metadatos XML/RDF, así como la facilidad para expresar clases de este último, unido a la proliferación de software, como por ejemplo Protegé (*Vid.* Capítulo 5, nota 220, Fig. 54) para la manipulación de *schemas* y vocabularios RDF, han hecho que proliferen proyectos para la creación de ontologías como sistema de recuperación de información semántica. Uno de los proyectos más notables es On-to-knowledge (1999-2002) <<http://www.ontoknowledge.org>> dentro del IST del V Programa Marco de la UE, cuyo primer resultado es la propuesta de un estándar específico para la creación de ontologías, denominado OIL, *Ontology Inference Layer*, compatible con RDFS. Asimismo, el W3C ha creado un grupo de trabajo (WebOnt) sobre el desarrollo de ontologías en el marco del proyecto global de Web semántica. Una visión muy interesante y completa de las ontologías, RDF y la nueva representación del conocimiento en Internet desde un punto de vista semiótico y lingüístico es la que da Sowa. *Vid.* John F. Sowa. *Ontology, Metadata and Semiotics* [documento HTML]. Best Web Internet, rev. 11 de septiembre de 2000. Disponible en: <http://www.bestweb.net/~sowa/peirce/ontometa.htm> (consultado el 12 de febrero de 2001). Una contribución notable para nuestra visión de las ontologías, desde la perspectiva de RDF como núcleo de la Web semántica, es el capítulo 8 de la obra de Hjelm, titulado *Reasoning about Metadata: Rules and Ontologies*. Johan Hjelm. *Op. cit.*, p. 201 y ss.

dependerá del nivel de normalización de los *schemas* y también de la estandarización y aplicación de *schemes*. No sólo es preciso normalizar las propiedades que evocan los metadatos, sino también sus valores⁹². Este aspecto es especialmente trascendente desde el punto de vista de que la mayoría de las búsquedas que se hacen tanto en Internet como en las bibliotecas, y por ello también en las bibliotecas digitales, son de naturaleza temática, donde la coherencia del vocabulario puede ser el problema central para la recuperación basada en metadatos. Así por ejemplo se reconoció en el taller sobre metadatos de Luxemburgo, al que ya hemos hecho alusión:

*[...] los estándares de metadatos, además de ser mecanismos de descripción de UNA materia o sector específico, son necesarios para asegurar la interoperabilidad en la recuperación de información en Internet*⁹³.

Afirmaciones como ésta plantean la necesidad de mecanismos de descripción controlados de la materia de los DLOs, que nos hace pensar no sólo en la normalización de estructuras sino también de la terminología necesaria para la descripción y recuperación en sistemas de información temáticos. Un componente crítico en el éxito de los metadatos es su creación y mantenimiento de forma cooperativa y distribuida. Esta fórmula para afrontar el crecimiento de los DLOs, bien en toda la Red o en un contexto abarcable de biblioteca digital, es una idea interesante pero tiene el peligro de que cada agente creador de metadatos asigne sus propias palabras clave, y que éstas no coincidan con las asignadas por el resto de los

⁹² Por ejemplo, no sólo es importante contar con un elemento `Subject` en el DC, un elemento `<textClass>` en la cabecera TEI, un campo `6XX`, en el formato MARC, etc., sino que los valores que se añadan a esos elementos han de ser tan consistentes como la propia estructura de la metainformación.

⁹³ *...metadata standards are necessary, in addition to sector-or subject-specific description mechanisms, to ensure interoperability in resource discovery on the Internet. Metadata Workshop... 1997. Op. cit., p.1. Las mayúsculas (UNA) son nuestras, para resaltar que el desarrollo de vocabularios controlados es aplicable a priori para un sistema de información de una sola materia, v. gr. una biblioteca virtual temática o subject gateway (Cfr. 8.2.2.).*

participantes en el proyecto. Los *schemes*⁹⁴ y tesauros mitigan este problema facilitando términos normalizados para indizar en virtud de las estructuras de metadatos, permitiendo una mayor coherencia y homogeneidad no sólo de estructuras de metainformación, sino también del contenido de las mismas. Las ventajas de la creación cooperativa de metadatos serán reales si éstos se crean de forma distribuida, pero con una entrada de datos normalizada. Hemos apuntado este aspecto tan importante de la asignación de metadatos al hablar de las aplicaciones para la creación de los mismos, donde calificábamos de ventaja la posibilidad de trabajar con *schemes* y tesauros como en el caso de la opción de edición y visualización de árboles de materia (*Tree Viewer*) de Metabrowser⁹⁵.

La normalización de los valores de los datos desde el punto de vista de las materias que tratan, servirá para dar coherencia a los metadatos que Kashyap y Sheth⁹⁶ denominaban *dependientes del contenido*. El contenido de los metadatos difiere desde la perspectiva de las normas de valor semántico que gobiernen su formulación⁹⁷ y en el sentido de que el entorno de la información electrónica distribuida en Internet es internacional en términos de vocabulario (todas las disciplinas) y lenguas (todos los idiomas). Todo ello conlleva un problema evidente de lenguaje, lógica y significado, mucho más amplio e intrincado de lo que podemos llegar a desarrollar dentro de los objetivos de esta investigación.

⁹⁴ Vid. Glosario, *schema*, *scheme*.

⁹⁵ Cfr. 4.3.2.2., Fig. 42.

⁹⁶ Virpul Kashyap, Amit Sheth. *Information Brokering... Op. cit.*, p. 19-22. Cfr. 3.5.

⁹⁷ En el caso de que las haya. Por ejemplo, en contextos muy concretos de recuperación de información en Internet, se utilizan normas de valor semántico como vocabularios o tesauros tanto para describir el contenido de los documentos como para realizar las búsquedas. V. gr. el sistema SOSIG (*Social Science Information Gateway*) <<http://www.sosig.ac.uk>> utiliza un tesoro derivado del *HASSET thesaurus*, desarrollado por el *Data Archive* en la Universidad de Essex partiendo del Tesoro de la UNESCO compilado por Jean Aitchison (Paris: Unesco, 1977). Cada vez es más frecuente también que los sistemas de recuperación de información de última generación se basen en ontologías y bases de conocimiento. Vid. 7.2., Oingo (Fig. 55).

La confrontación entre lenguaje controlado y lenguaje natural o texto libre (si son palabras directamente extraídas del propio texto) como principios para describir el contenido en la indización, son una constante en toda la bibliografía sobre recuperación de información (Lancaster, Blair, Rowley etc.). Podríamos basarnos en diversas aproximaciones a ambas perspectivas del análisis del contenido documental. Sin embargo, vamos a partir de las ventajas e inconvenientes del control del vocabulario aplicado a la recuperación de información en OPACs que señalan Aluri, Kemp y Boll⁹⁸, con el objetivo de diferenciar los matices del nuevo paradigma de información distribuido y heterogéneo de la Web, lo que nos lleva a replantear los principios bibliotecarios tradicionales.

Dentro de las ventajas genéricas de los lenguajes controlados que señalan dichos autores en el contexto de la recuperación por materias en OPACs, se puede destacar, matizadas por el contexto general de sistemas de recuperación de información en Internet⁹⁹ como anunciamos, las siguientes:

- La probabilidad de que el indizador y el usuario expresen el mismo concepto de la misma manera, de tal forma que se incrementan los resultados en el proceso de emparejamiento en la búsqueda. Esta afirmación, que incluso en el contexto de la recuperación en OPACs es incierta en muchos casos, en el contexto general de la recuperación de información en Internet se reduce aún más: en primer lugar, porque en la mayoría de los casos la indización se hace en texto libre y de forma automática; y en segundo lugar, en el caso de que el elemento de metadatos relativo a la materia estuviese normalizado según un tesoro, la propia globalidad de la Red conduce a que usuario no conozca el vocabulario o tesoro que pudo

⁹⁸ Rao Aluri, Aldasdair Kemp and John J. Boll. *Subject Analysis in Online Catalogs*. Englewood: Libraries Unlimited, 1991, p. 33 y ss.

⁹⁹ *Cfr.* Capítulo 7.

utilizar el autor de la indización, bien por disparidad idiomática, bien por imposibilidad de acceso a tal lenguaje controlado.

- Asegura la consistencia en la descripción de información, desde el punto de vista de que distintos analistas o el mismo indizador en dos momentos distintos, pueden asignar los mismos términos de indización. Sobre este aspecto existen muchas opiniones en contra en el contexto de la indización humana tradicional. Cleverdon, por ejemplo, estudió que incluso el mismo indizador en distintos momentos determina distintos términos de indización. En el contexto de los metadatos asignados a un DLO, si partimos del principio de que sean los propios autores/editores de los documentos electrónicos los que asignen los metadatos y también completen el contenido del elemento correspondiente a las materias (Keywords, DC.Subject, textClass, etc.), esta consistencia sería insignificante teniendo en cuenta el número de creadores/editores e indizadores potenciales de la Web.
- Incrementa la probabilidad de que el usuario y el indizador lleguen a la materia concreta que desean, gracias a la propia estructura combinatoria y a las relaciones que se establecen. En la información Web, para que las relaciones semánticas de un tesoro cumplan su función de especificidad, asociación, o generalización, tendría que estar accesible tanto en el *input* como en el *output* y ser parte integrante del proceso de creación de metadatos así como del proceso de búsqueda; .
- Contribuye a la velocidad en el proceso de búsqueda porque permite al usuario averiguar rápidamente que el ítem que busca no está accesible, ya que el lenguaje controlado garantiza que no sea necesario buscar por un gran número de sinónimos. Esta garantía de velocidad en la búsqueda, que también daría lugar a mucha discusión en el contexto de los OPACs, tampoco es cierta si pensamos en sistemas de búsqueda genéricos en Internet de indización post-coordinada. A pesar de que indizasen la etiqueta <META> al indizar también el texto completo, casi siempre devolverán resultados ante un término supuestamente controlado

desde el punto de vista de un usuario, que no haya sido asignado en las etiquetas <META> correspondientes con la misma intención.

Lancaster¹⁰⁰ resume las ventajas de los lenguajes controlados en los sistemas de recuperación en tres funciones fundamentales: *tiende a reducir las ambigüedades semánticas, mejora la consistencia en la representación de la materia y facilita la realización de búsquedas amplias*. Los argumentos tradicionales en contra del uso de lenguajes controlados que, en principio, serían las ventajas del lenguaje natural, entendidos o enunciados en el sentido tradicional nos llevarían a la defensa de los sistemas de indización *full-text*. Por ello, al pensar en las ventajas del lenguaje natural (o inconvenientes del lenguaje controlado) tendremos que partir del supuesto de una indización y búsqueda por materias basada en la estructura de metadatos, completando el elemento correspondiente a la materia o palabras clave con lenguaje libre o natural asignado por el creador del registro de metadatos. De lo contrario, si pensamos en una mera indización en texto libre en sentido estricto, estaríamos defendiendo los sistemas de búsqueda *full-text* en los que se basan la mayoría de los sistemas de recuperación genéricos de Internet, que soslayaremos en el capítulo siguiente. Mencionada esta salvedad de interpretación, señalaremos algunos de los inconvenientes de los lenguajes controlados:

- El lenguaje controlado está limitado, por definición, a unos cuantos miles de palabras y a un número limitado de lenguas (en el caso de que sea un tesoro multilingüe).
- Lancaster¹⁰¹ destaca, además, que los sistemas de lenguaje natural ofrecen una ventaja sobre los sistemas que utilizan el lenguaje controlado ya que el uso de un lenguaje ilimitado (lenguaje libre) permite una gran especificidad

¹⁰⁰ Frederick W. Lancaster. *El control del vocabulario en la recuperación de información*. Valencia: Universitat, 1995, p. 178-179.

¹⁰¹ *Ibid.*, p. 178.

en la recuperación, aunque esa especificidad se torne ambigüedad en algunos casos de búsqueda global.

- El lenguaje natural es adecuado para la indización y recuperación en comunicaciones científicas específicas. Sin embargo, y a pesar de la univocidad entre significante y significado en el signo lingüístico científico, la descripción en lenguaje natural (en la etiqueta destinada a las materias en un formato de metadatos) sólo sería útil si los registros de metadatos formasen parte de una biblioteca digital concreta, de un servicio de búsqueda específico para una comunidad de usuarios¹⁰².
- *Los lenguajes controlados altamente estructurados están obsoletos para la indización y recuperación.* Esta afirmación sólo es parcialmente cierta, ya que para representar semánticamente la jerarquización inherente de los recursos de Internet pueden resultar útiles las relaciones verticales que se establecen en lenguajes controlados como los tesauros, máxime si pensamos en la potencialidad de las ontologías.
- Por otra parte, se ha señalado también que los lenguajes controlados, necesitan un conocimiento previo y que su utilización para la indización y búsqueda por materias está confinada a profesionales de la información, no al usuario final.

A pesar de todo esto —y de otras múltiples interpretaciones que podríamos traer aquí como representación de la confrontación secular de ambos métodos de

¹⁰² Por ejemplo, describir los documentos médicos con el propio lenguaje natural (científico-técnico de la Medicina) sería suficiente en una colección virtual como la de OMNI *Cfr.* (8.2.2.). Sin embargo, si se pretende realizar una búsqueda genérica por materias en un sistema global, el lenguaje natural científico pierde univocidad (*v.gr.* el término aborto o análisis, que en el contexto médico no dan lugar a polisemia, si se hiciese una búsqueda en toda la Web por palabra clave en un buscador genérico, se obtendrían resultados subjetiva o semánticamente poco relevantes).

indización y búsqueda basados en el lenguaje—, la práctica bibliotecaria¹⁰³ evidencia la necesidad de lenguajes controlados para optimizar la recuperación de información. Lo difícil ahora es determinar qué sistemas de representación y organización del conocimiento se necesitan en un contexto de información distribuido como el de Internet y en qué medida son aplicables por el propio creador/autor del documento, que lo será también de los metadatos. Si bien es cierto que indudablemente podemos aplicar muchos de los sistemas utilizados en las bibliotecas tradicionales¹⁰⁴, también es obvio, según lo que hemos comentado en este trabajo, que existen otros desafíos en el nuevo entorno multilingüe, multicultural e interdisciplinar de la World Wide Web que requieren nuevos planteamientos y soluciones.

Por un lado, la aceptación de la importancia de los vocabularios controlados y los esquemas formales de clasificación para la catalogación o la descripción (utilizando un término más aséptico y menos vinculado al mundo bibliotecario tradicional) de los recursos de la Web está cada vez más difundida. Sin embargo, la total diversidad de información de la Web, ha puesto de manifiesto en los distintos

¹⁰³ Las bases de datos tradicionales, tanto en línea como en CD-ROM, mantienen sus tesauros y sistemas de clasificación, p. ej. MeSH, ERIC, INSPEC, etc. creados para sistemas específicos de indización y resumen. La búsqueda por términos basada en estos lenguajes controlados (descriptores del tesoro), se ve enriquecida o complementada por la búsqueda en texto libre del resumen. Lo que nos hace pensar que la búsqueda en lenguaje natural y la búsqueda según un vocabulario controlado, no son técnicas antagónicas ni excluyentes en la recuperación de información en las bibliotecas tradicionales y tampoco deben serlo en las bibliotecas digitales. De hecho son muchas las investigaciones en recuperación de información que concluyen la complementariedad de ambas. V. *gr* Muddamalle, se refiere a ambas como "una combinación ideal". Manikya Rao Muddamalle. Natural Language versus Controlled Vocabulary in Information Retrieval: A Case Study in Soil Mechanics. *Journal of the American Society for Information Science*, August 1998, vol. 49, n° 10, p. 887.

¹⁰⁴ En este sentido nos referimos a la utilización de vocabularios controlados tradicionales y preexistentes, como por ejemplo la LCSH. Según Pat Oddy —*Op. cit.*, p. 163— y otros bibliotecarios del dominio lingüístico anglosajón, la lista de encabezamientos de materia de la Biblioteca del Congreso de Washington, se ha convertido en el estándar internacional *de facto* para la indización y el acceso por materias en inglés. No en vano, en algunos software para la creación de metadatos, como por ejemplo *Reggie*, según hemos comentado anteriormente (*Cfr.* 4.3.2.1.), incluyen en el elemento "Subject" de la plantilla de edición de metadatos según el modelo de la DCMI, la posibilidad de añadir como subelemento "LCSH", para indicar que la descripción por materias del objeto electrónico en cuestión, se ha realizado según este vocabulario.

sistemas de búsqueda las limitaciones de las taxonomías y clasificaciones existentes para organizar el conocimiento humano. Ni las taxonomías constituidas *ad hoc* en el contexto de los nuevos directorios de información de la Web como Yahoo <<http://www.yahoo.com>>, ni las clasificaciones bibliotecarias tradicionales utilizadas por otros sistemas de recuperación de cariz bibliotecario¹⁰⁵, como CyberStacks <<http://www.public.iastate.edu/~CYBERSTACKS/homepage.html>> pueden asumir la tarea, aparentemente imposible según nuestra opinión, de modelar y representar la vasta naturaleza del conocimiento humano que, en mayor o menor medida, se encuentra representado en la Web.

Señalamos a continuación algunos aspectos que se pueden discutir con relación a la creación de lenguajes controlados y ficheros de autoridad para unificar el contenido de los metadatos, de tal forma que la recuperación de información sea más consistente y eficaz.

- En primer lugar, de forma intuitiva cabría pensar que la solución más acertada desde el punto de vista documental sería crear un macrotesaruro pluridisciplinar y multilingüe que sirviera para que los creadores de los metadatos asignaran las etiquetas relativas a la descripción característica de las materias de los objetos digitales de información (DLOs) en el momento de su creación para asegurar una recuperación general consistente. Esta solución es desmedida e inviable en términos económicos y de consenso internacional normalizador, teniendo en cuenta, además, que la producción de contenidos en la Red responde a una motivación más individual o individualista, que hace que trascienda la posibilidad de cumplir una norma de vocabulario así planteada.

¹⁰⁵ Cfr. 7.2.

- Por otra parte Constança Espelt¹⁰⁶ señala que el contenido de los metadatos puede cubrir dos tipos de datos: por un lado, de concepto, dirigidos hacia la categorización conceptual, y por otro, de forma, que incluyan aspectos relativos al lenguaje, localización geográfica, tipo de medio, audiencia y criterios de calidad. Para esto son necesarios ficheros de autoridad que permitan una única representación de contenido de los metadatos tanto relativos a la representación de las materias como de la forma¹⁰⁷. Sin embargo, como señala Miller¹⁰⁸, los esfuerzos para desarrollar listas controladas de términos (independientemente de que sirvan para controlar el contenido conceptual o formal de los metadatos) tienen lugar en disciplinas relativamente aisladas e individuales, de tal forma que corroboramos la existencia de bibliotecas digitales asociadas a campos temáticos concretos, en las que la recuperación de información puede llevarse a cabo siguiendo principios bibliotecarios incluso en lo que se refiere a control del vocabulario. Para que el acceso a estas colecciones individuales se universalice, es preciso conseguir también una interoperabilidad entre sistemas terminológicos, integrando diversas ontologías y terminologías controladas y facilitando el acceso a ellas¹⁰⁹.

¹⁰⁶ Constança Espelt. Improving Subject Retrieval: User-friendly Interfaces and Effectiveness [documento HTML]. *bid: Biblioteconomia i Documentació*, juny 1998, nº 1. Disponible en: <http://www.ub.es/div5/biblio/bid/espe198.htm> (consultado el 25 de junio de 1998).

¹⁰⁷ En este sentido, *Cfr.* Capítulo 3, nota 67. *Draft Thesaurus of Information Object... Op. cit.*, http://alexandria.sdc.ucsb.edu/~lhill/MultiTes_files/objtype.tes/tagged.txt

¹⁰⁸ Paul Miller. Controlled Vocabulary [presentación PPT]. En: *Metadata Workshop... Op. cit.*, <ftp://ftp.echo.lu/pub/libraries/metadata/miller.ppt>

¹⁰⁹ *Cfr.* 8.3.2. donde hablaremos de la interoperabilidad también desde el punto de vista semántico.

- Otro de los aspectos que hay que tener en cuenta, confirmado por la apreciación de Berners-Lee¹¹⁰, es que *sería ingenuo imaginar que los metadatos de cada recurso usarán un solo vocabulario* (tanto de etiquetas que designen las propiedades de los recursos, como de los valores de esas propiedades). Según esto, desde el punto de vista de los vocabularios para el contenido, necesitarán convivir varios tipos de ellos (tesauros, ontologías, diccionarios geográficos, etc.) y de esquemas de clasificación en el mismo espacio de información o en el mismo registro de metadatos. Esto implica que los ordenadores tienen que ser capaces de identificar la fuente de autoridad para los términos o marcas de clase, y por consiguiente, en las descripciones de metadatos se declararán *namespaces* relativos tanto a *schemas* como a *schemes*, para asegurar que las etiquetas usadas para identificar las diferentes autoridades son únicas y carecen de ambigüedad.

En estos aspectos que acabamos de subrayar, así como en las reflexiones anteriores seleccionadas de la diatriba tradicional en Documentación entre lenguajes controlados e indización en texto libre, se aprecia una corroboración más de nuestra tesis de la funcionalidad de los metadatos a nivel local o finito, en bibliotecas digitales particulares, apoyada en este caso en el control de un vocabulario científico de una materia en particular. Por otra parte, de todo este bosquejo del papel de los vocabularios controlados y listas de autoridad para determinar el contenido de los metadatos, se desprende la necesidad de redefinir el concepto de herramientas terminológicas en un contexto de información distribuido. Ni los lenguajes postcoordinados, ni las clasificaciones pueden reducir, en gran medida, el problema del contexto de la información en la Red, señalado por Ward y Wood¹¹¹, y tampoco son suficientes para categorizar la inabarcable variedad de información de la Red. En

¹¹⁰ Tim Berners-Lee. Metadata Architecture... *Op. cit.*, <http://www.w3.org/DesignIssues/Metadata.html>

¹¹¹ Nigel Ward and Andrew Wood. *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA98>

este sentido, nos sumamos a la opinión de Carol Mandel¹¹², quien considera que todavía no se han resuelto los problemas de la búsqueda por materias en los OPAC, a lo que añadimos que más lejos todavía está la resolución del problema de la búsqueda por materias global en Internet, ya que se trata de un medio multilingüe y heterogéneo, donde la cantidad de información es inabarcable. Sin embargo, utilizar *schemes* de contenido para facilitar la estructuración, la gestión y el proceso de los datos basados en el conocimiento y un acceso sistemático a dichas estructuras de conocimiento (tesauros, ontologías, sistemas de clasificación, bases de datos léxicos, taxonomías, etc.) es factible aplicándolos a colecciones individuales o bibliotecas digitales de una materia concreta. Así, por ejemplo, Traugott Koch¹¹³ recopila más de cien tesauros en Internet para completar el elemento DC . Subject del Dublin Core, según distintos dominios temáticos. Aunque debemos señalar que *todavía no está claro si la utilización de estos tesauros contribuye realmente a la descripción y recuperación de recursos electrónicos*¹¹⁴.

Además de estas cuestiones elementales basadas en la práctica bibliotecaria tradicional, la normalización de sistemas de organización del conocimiento para la Web y para las bibliotecas digitales está cambiando. Son muchos los autores que reconocen esta realidad y que proponen nuevas formas de normalizar el valor de los datos a través de Sistemas de Referencia Terminológica (SRT¹¹⁵) o Sistemas y

¹¹² Carol Mandel. *Change and Continuity in Subject Authority Control* [documento HTML]. OCLC, rev. 22 de mayo de 1998. Disponible en: <http://www.oclc.org/oclc/man/9391ausy/mandle.htm> (consultado el 30 de junio de 2000).

¹¹³ Traugott Koch. *Controlled Vocabularies, Thesauri and Classification Systems Available in the WWW. DC Subject* [documento HTML]. Lund: University Libraries, rev. 27 de septiembre de 2001. Disponible en: <http://www.lub.lu.se/metadata/subject-help.html> (consultado el 13 de octubre de 2001).

¹¹⁴ Ali Asghar Shiri, and Crawford Revie. Thesauri on the Web: Current Developments and Trends. *Online Information Review*, 2000, vol. 24, nº 2, p. 273-279

¹¹⁵ Larry Fitzwater and Linda Spencer. EPA Terminology Reference System (TRS) [presentación PPT]. En: *CENDI Conference Controlled Vocabulary and The Internet (1999. Bethesda)*. CENDI, [cont.]

Servicios de Organización del Conocimiento en Red (NKOS¹¹⁶). El uso de estas herramientas de descripción del contenido debe llevarse a cabo por programas de software que automáticamente transformen las preguntas de los usuarios en los sinónimos incluidos en el tesoro, o incluso la traducción de los términos de la pregunta en lenguajes alternativos, o mapeando entre distintos esquemas de clasificación y autoridades terminológicas. Las nuevas bibliotecas digitales deben prestar un *acceso por materias alterno*¹¹⁷, que implica proporcionar más de una orientación de vocabulario, de tal forma que las nuevas bibliotecas virtuales sean útiles a diferentes usuarios, a través de la indización y clasificación, utilizando múltiples esquemas (*shemes*) o intercambiando información entre *schemes* alternativos.

Esto nos lleva a pensar en la necesidad de desarrollar interfaces inteligentes diseñadas para ayudar a buscar en múltiples vocabularios, ya sea mediante algoritmos de comparación o mediante pantallas de ayuda que sugieren alternativas en el vocabulario, lo que Jessica Milstead¹¹⁸ denomina *tesauros dirigidos al usuario final* o *redes semánticas*, que se diferencian de los tesauros tradicionales básicamente en la *inclusión y organización de los términos y en su presentación*. Bajo esta visión, los tesauros reflejan y organizan el vocabulario total especializado de los usuarios de un

Subject Analysis and Retrieval Working Group, 1999. Disponible en: <http://www.dtic.mil/cendi/presentations/Spencer-Fitzawater.ppt> (consultado el 27 de agosto de 2001).

¹¹⁶ NKOS (*Networked Knowledge Organization Systems/Services*) es una iniciativa creada en 1997, con el objetivo de promocionar la discusión para crear un modelo funcional y de datos que permita un Sistema de Organización del Conocimiento, esto es, sistemas de clasificación, tesauros, diccionarios geográficos y ontologías, como servicios de información interactiva en red, de tal forma que soporten la descripción y recuperación de recursos de información diversos a través de Internet. Vid. <http://nkos.slis.kent.edu>

¹¹⁷ Gail Hodge. *System of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington: Council on Library and Information Resources, April 2000, p. 17.

¹¹⁸ Jessica L. Milstead. Use of Thesauri in Full-Text Environment [documento HTML]. Disponible en: <http://www.jelem.com/useof.htm> (consultado el 5 de julio de 2001).

dominio de conocimiento. Es lo que Soergel¹¹⁹, de manera ideal, define como la formalización deseada del valor de los datos, como *tesauro virtual* que proporcione un acceso transparente a múltiples bases de conocimiento (tesauros, diccionarios y otros recursos léxicos).

Las normas para la construcción de tesauros en la organización de la información tradicional están perfectamente definidas y aplicadas a través de estándares internacionales (ISO 2788) o nacionales (ANSI/NISO Z39.19-1993, UNE-50-106-90). Sin embargo, la necesidad de integración de las herramientas de control del valor de los datos, con las definiciones de estructuras de propiedades (metadatos), hace que se estén buscando nuevos estándares para la creación de tesauros electrónicos que se integren en el nuevo paradigma de las bibliotecas digitales y que respondan a un planteamiento realista de recuperación de información distribuida entre ellas.

La búsqueda de nuevas normas para las nuevas necesidades de organización de la información es un aspecto recurrente en este trabajo, en consonancia con el crecimiento de la información electrónica y la necesidad de replantear los viejos instrumentos de organización de la información. Ante el crecimiento individual de tesauros y herramientas terminológicas basadas en la Web, son muchos los autores que han reconocido la necesidad de un estándar para la creación de ese nuevo tipo de tesauros. Con esa finalidad surgió en 1997 NKOS, e igualmente Hodge, Soergel, Shiri y Revie ponen de manifiesto esta nueva visión de las normas para la creación de sistemas de organización del conocimiento en red¹²⁰. Así, en noviembre de 1999 se

¹¹⁹ Dagobert Soergel. Enriched Thesauri as Networked Knowledge Bases for People and Machines [documento PDF]. En: *CENDI Conference Controlled Vocabulary and The Internet (1999. Bethesda)*. CENDI, Subject Analysis and Retrieval Working Group, 1999. Disponible en: <http://www.dtic.mil/cendi/presentations/cendisoergel.pdf> (consultado el 27 de agosto de 2001).

¹²⁰ NKOS <<http://nkos.slis.kent.edu>>; Gail Hodge. *Op. cit.*, p. 31; Dagobert Soergel. *Op. cit.*, <http://www.dtic.mil/cendi/presentations/cendisoergel.pdf>; Ali Asghar Shiri and Carwford Revie. *Op. cit.* p. 278.

celebró, bajo los auspicios de la NISO, de la Sociedad Americana de Indizadores (ASI) y de la Asociación para las Colecciones de Bibliotecas y Servicios Técnicos (ALCTS), el taller *Tesoros electrónicos: planificando un estándar*¹²¹ con la intención de abordar el problema de la normalización de tesauros para la recuperación de información electrónica. Se detectó la carencia de normas para publicar tesauros en la WWW que sirvan para fundamentar la recuperación en ella. Según este taller, un estándar para los tesauros electrónicos debe contemplar: a) los criterios y/o métodos para la generación de tesauros con medios automatizados; b) las relaciones semánticas entre los términos como ayudas para el análisis y recuperación de información; c) la presentación de una amplia variedad de tesauros; y d) los protocolos de interoperabilidad, las estructuras y/o la semántica aplicable a los tesauros.

Con una finalidad semejante a ésta última, reflexionar en torno a la normalización e interoperabilidad terminológica, se celebró en el seno de MODELS¹²², otro taller a principios del año 2000. Este taller destacó el papel tradicional de la terminología y de los tesauros en la recuperación de información, pero haciendo hincapié en el acceso *online* creciente a los recursos de información, por usuarios no especializados, que implica la necesidad de crear una lista de términos de alto nivel para la un acceso por materias interdisciplinar, a los recursos electrónicos.

Tanto XML como RDF son las propuestas para normalizar, no sólo la estructuración de metadatos orientados a la descripción de recursos, sino también la

¹²¹ *CENDI Conference Controlled Vocabulary and The Internet*: http://www.dtic.mil/cendi/proj_cont_vocab.html

¹²² Para consultar los objetivos y los temas que se presentaron en dicho taller sobre terminología. Vid. MODELS (MOVing to Distributed Environments for Library Services). *Terminology Workshop*: <http://www.ukoln.ac.uk/dlis/models/models11> Este *workshop* profundizaría en el desarrollo del proyecto HILT (*High-Level Thesaurus*) <<http://hilt.cdrl.strath.ac.uk>>, del que tendremos oportunidad de hablar al tratar la interoperabilidad del acceso por materias. Cfr. 8.3.2.

estructuración de metadatos orientados a las materias, esto es, de las herramientas terminológicas que se integren en la recuperación de información. En dicho taller para la reflexión en torno a los estándares para la creación vocabularios controlados, se reconoció que los formatos de metadatos existentes hacen poco uso de los tesauros, quizás justamente por la escasa integración de estos sistemas de organización del conocimiento en la nueva estructuración de la información electrónica. A raíz de este *workshop* y de las actividades desarrolladas en el seno de NKOS, han surgido distintas iniciativas de representación de tesauros a través de RDF, especificando distintos tipos de relaciones (SN, UF, BT, NT, USE, etc.) como tipos de propiedades. Tal es el caso de CERES¹²³ que ha desarrollado un esquema en RDF para la definición de tesauros.

EJEMPLO: Término "biosfera" del tesoro CERES representado en RDF

```
<RDF:RDF>
xmlns="http://ceres.ca.gov/thesaurus/RDF/"
xmlns:rdf="http://www.w3.org/TR/WD-rdf-syntax/"
<SUBSET rdf:resource="http://localhost/thesauri/Theme?biosphere">
  <THESAURUS resource="http://localhost/thesauri/Theme">
    <DESCRIPTOR ID="Biosphere" LABEL="Biosphere" SN="Organisms and
ecosystems">
      <CAT LABEL="Natural environment" RDF:RESOURCE="Natural
environment"/>
      <TT LABEL="Earth" RDF:RESOURCE="Earth"/>
      <BT LABEL="Earth" RDF:RESOURCE="Earth"/>
      <RT LABEL="Biosphere [GEMET]" RDF:RESOURCE="Biosphere [GEMET]"/>
      <RT LABEL="Biosphere [LCSH]" RDF:RESOURCE="Biosphere [LCSH]"/>
      <NT LABEL="Organisms" RDF:RESOURCE="Organisms"/>
      <NT LABEL="Biodiversity" RDF:RESOURCE="Biodiversity"/>
      <NT LABEL="Biological processes" RDF:RESOURCE="Biological
processes"/>
      <NT LABEL="Ecosystems" RDF:RESOURCE="Ecosystems"/>
      <NT LABEL="Evolution" RDF:RESOURCE="Evolution"/>
    </DESCRIPTOR>
  </THESAURUS>
</SUBSET>
</RDF:RDF>
```

¹²³ CERES/NBII. *The RDF Thesaurus Description Standard* [documento HTML]. California: Environmental Resources Evaluation System, California Resources Agency, rev. 26 de septiembre de 2000. Disponible en: <http://ceres.ca.gov/thesaurus/RDF.html> (consultado el 28 de septiembre de 2000).

En el seno del ILRT (*Institute for Learning and Research Technologies*)¹²⁴ también se está gestando una especificación para la creación de tesauros en RDF surgida a mediados del año 2000, donde se describe la aplicación de RDF a la construcción de tesauros, entendiendo éstos como estructuras de datos. Tanto el proyecto de CERES como la iniciativa embrionaria del ILRT parten de que RDF, como hemos señalado en el capítulo 5, es un lenguaje potente para protagonizar cualquier tipo de descripción y, además, está basado en XML; así pues, el modelo de datos de RDF junto con un esquema (*schema*) bien definido y un lenguaje de interrogación, puede representar una visión Web de conceptos y términos interrelacionados de uno o más tesauros¹²⁵.

Con objetivos semejantes, se presentó, en noviembre de 2000, el *Vocabulary Mark-Up Language* (Voc-ML¹²⁶) para definir la estructura de contenido de los sistemas de organización del conocimiento. Es una propuesta para la representación de *schemes* que contribuya a la creación de la Web Semántica. Esta DTD de XML, que acepta la DCMI, proporciona etiquetas y sintaxis para identificar cada término y sus relaciones con otros términos según se establece en la norma Z39.19. Se trata de constituir una DTD lo más genérica posible, que permita trabajar con distintos sistemas de organización del conocimiento (ficheros de autoridad y tesauros jerárquicos, esquemas de clasificación y listas de encabezamientos de materia). También debemos señalar iniciativas particulares de normalización de creación de vocabularios, por ejemplo, el TDDL (*Terminology Definition Description*

¹²⁴ Phil Cross, Dan Brickley, Traugott Koch. *RDF Thesaurus Specification (draft)* [documento XHTML]. ILRT, 24 de enero de 2000, rev. 24 de enero de 2001. Disponible en: <http://ilrt.org/discovery/2001/01/rdf-thes> (consultado el 26 de agosto de 2001).

¹²⁵ Nótese como en el ejemplo anterior el término biosfera aparece relacionado en otros dos lenguajes controlados: la lista de encabezamientos de materia de la *Library of Congress* (LCSH) y el *General Multilingual Environmental Thesaurus* (GEMET).

¹²⁶ *Vocabulary ML: Metacode Strawman DTD* [documento RTF]. NKOS, University of Kent, 11 de octubre de 2000. Disponible en: <http://nkos.slis.kent.edu/VOCML-1.DOC> (consultado el 20 de noviembre de 2000).

Language), surgido en el seno del Open Healthcare Group como un borrador de un lenguaje dirigido a solucionar también el problema de la terminología. En este caso TDDL supone una extensión de XHTML para describir terminologías. Para ello, incorpora conceptos del RDFS, así como su sintaxis y funcionalidad¹²⁷.

Como venimos diciendo a lo largo de este capítulo, las guías, procedimientos y estándares mutuamente establecidos se convierten en políticas que ayudan a definir el ámbito de las interacciones entre las organizaciones que cooperan. Sin embargo, queda mucho por hacer en términos de normalización, tanto desde el punto de vista de la estructura como de la semántica y mucho más aún desde el punto de vista del control terminológico del valor de los datos. La normalización en este sentido demuestra la solvencia de XML y RDF como estándares en los que fundar la Web Semántica. Es necesario, además, un compromiso por parte de los actores encargados del desarrollo de DLOs para la Web, así como de los gestores de la información implicados en la creación de bibliotecas digitales, para la utilización de estas recomendaciones o estándares, independientemente de que su adecuación esté validada por la ISO o no. Para que la recuperación eficaz y relevante basada en sistemas de metadatos trascienda a un dominio informativo determinado y a bibliotecas digitales creadas como esfuerzos particulares, es preciso incentivar la interoperabilidad tanto desde el punto de vista de las estructuras semánticas de los datos, como de los valores de los metadatos. Retomaremos, pues, el tema de la normalización al tratar el problema del acceso universal a las bibliotecas digitales.

¹²⁷ Vid. Jonathan Borden. *Terminology Definition Description Language (TDDL) 1.0* [documento XHTML]. Open Healthcare Group, rev. 23 de marzo de 2001. Disponible en: <http://www.openhealth.org/RDDL/tddl> (consultado el 10 de agosto de 2001). Los elementos del esquema TDDL están basados en el esquema RDF, así por ejemplo una clase en TDDL (`tddl:Class`) se define a partir de una clase del RDFS (`rdfs:Class`) etc., y se puede transformar a un esquema RDF a través de XSLT <http://www.openhealth.org/RDDL/tddl2rdfs.xsl>.

PARTE III:
**Acceso global a la información: recuperación, metadatos y
bibliotecas digitales**

7. Metadatos y recuperación de información en Internet. 7.1. Recuperación de información en la Red: consideraciones generales. 7.2. Tendencias de los sistemas de búsqueda... ¿metadatos? 7.3. Búsqueda en texto completo *versus* búsqueda en metadatos: ventajas e inconvenientes. **8. Bibliotecas digitales y metadatos.**
8.1. ¿Biblioteca digital/virtual? La biblioteca como concepto y la diversidad designativa 8.2. Bibliotecas virtuales: del oxímoron del término a la realidad práctica.
8.3. Bibliotecas digitales y globalización.

CAPÍTULO 7

METADATOS Y RECUPERACIÓN DE INFORMACIÓN EN INTERNET

El problema del almacenamiento y recuperación de información llama cada vez más la atención. Es simple, tenemos una inmensa cantidad de información cuyo acceso rápido y preciso es cada vez más difícil [...] Con la llegada de los ordenadores, se desarrollaron muchas ideas para ofrecer sistemas de recuperación rápidos e inteligentes. En las bibliotecas [...] los ordenadores han asumido de manera satisfactoria algunas de las tareas más simples, como la catalogación y la administración general. Sin embargo, el problema de la recuperación efectiva continúa sin resolverse¹.

Esta reflexión de Rijsbergen, a pesar de estar escrita hace treinta años, podría resumir perfectamente, uno de los problemas de fondo que subyace a esta investigación: *el problema de la recuperación efectiva continúa sin resolverse*. La informatización de las unidades de información, la evolución de los OPACs, la automatización de los catálogos gracias al formato MARC, y todo un universo teórico de investigaciones y evaluaciones de las formas y procesos de búsqueda y recuperación, han contribuido a mitigar el problema de la recuperación de información (RI) en las bibliotecas y centros de documentación. Ahora, la realidad que nos ocupa en este trabajo es distinta: el documento electrónico y el desarrollo de Internet, han provocado que ahora sí sea posible el acceso rápido a la información, sin embargo el problema de fondo continúa siendo el mismo que en los 70s de Rijsbergen: la recuperación sigue sin ser precisa, con el inconveniente añadido de la

¹ C. J. van Rijsbergen. *Op. cit.*, <http://www.dcs.glasgow.ac.uk/Keith/Preface.html>

complicación de las tareas de descripción y análisis del conocimiento de los documentos distribuidos en la Red.

Por otra parte, Bates² argumenta que la evaluación de los sistemas de recuperación de información continuará siendo problemática teniendo en cuenta que: implican un procesamiento cognitivo del lenguaje, conllevan un serio problema de escalabilidad y que las necesidades de los usuarios varían con el tiempo, con el dominio temático y con otros factores que reclaman mecanismos especiales de indización y recuperación. Esta última circunstancia de la vulnerabilidad o mutabilidad de las necesidades de los usuarios es patente en el entorno de la RI en Internet y, desde nuestro punto de vista, es lo que ha llevado a desarrollar metadatos con propósitos específicos y también bibliotecas digitales temáticas, así como un sinfín de propuestas, soluciones, iniciativas y un sinfín de "mecanismos especiales", para intentar resolver ese problema.

Auque en la II parte de este trabajo, hemos apuntado diversos modelos y estándares que ofrecen cierta esperanza para la descripción estructurada de documentos entendidos como objetos a fin de mejorar la recuperación, todos esos modelos pueden ser inútiles si no se aplican de forma generalizada. Por ello, en esta parte de la investigación (III) trataremos de analizar la utilidad de los metadatos diferenciando dos contextos diferentes: en este capítulo, el entorno de la recuperación de información global en Internet donde estudiaremos la proyección de los formatos de metadatos de carácter general (etiqueta <META> en HTML y la semántica del Dublin Core) en los Sistemas de Recuperación de Información de Internet (SRII); y en el capítulo siguiente, las bibliotecas digitales o contextos finitos, locales y abarcables de información, así como el nivel de aplicación de los formatos de propósito específico en los que se fundamentan.

² M Bates. Indexing and Access for Digital Libraries and the Internet: Human, Database and Domain Factors. *Journal of American Society for Information Science*, 1998, vol. 49, nº 12, 1185-1205. Citado en. G. G. Chowdhury. *Introduction to Modern Information Retrieval*. London: Library Association Publishing, 1999, p. 423.

Hecha esta precisión, corresponde ahora plantear una aproximación a los sistemas de búsqueda globales o genéricos (SRII) pero centrando nuestro análisis en el uso que éstos hacen de la metainformación; ya que plantear aproximaciones más amplias al tema de la RI en Internet trasciende los objetivos de esta investigación³. A lo largo de este capítulo analizaremos pues, el entorno de los SRII y su relación con los metadatos, atendiendo a tres aspectos fundamentales: en primer lugar, las características generales de la recuperación de información en la Red como contexto básico de la búsqueda, en segundo lugar, aportaremos una reflexión sobre las tendencias de dichos sistemas y el uso que hacen de los metadatos y por último, compararemos la recuperación basada en el texto completo, predominante en la Web, frente a la recuperación basada en estructuras de metadatos, como prospectiva ideal de lo que debería ser la búsqueda de información en la Red, basada en estructuras

³ El tema de la búsqueda y recuperación de información en Internet se puede tratar desde múltiples puntos de vista que van desde el cómo buscar y sacar el mayor partido a las herramientas genéricas de recuperación de información en la Red, a análisis descriptivos de cada uno de los servicios de búsqueda en Internet, prestando especial atención a su evolución y aspectos distintivos. Así, este tema ha suscitado múltiples publicaciones, inclusive en español, amén de un sinfín de recursos disponibles en la Web sobre el tema. *Vid.* Bibliografía.

Una excelente revisión del *statu quo* de la búsqueda en Internet es la que realiza por ejemplo, el profesor Hernández partiendo de las diferencias de Internet con las bases de datos tradicionales. Además explica la realidad de la búsqueda en Internet tanto desde las estrategias, como desde el tipo de información que se puede encontrar en la Red (software, personas, páginas Web), etc. así como las tecnologías utilizadas para ello. Antonio Hernández Pérez. La búsqueda y recuperación de información en Internet. En: *La Sociedad de la Información: política, tecnología e industria de los contenidos*. Mercedes Caridad Sebastián, coord. Madrid: Fundación Ramón Areces, 1999, p. 213-238. Una de las últimas publicaciones monográficas y de carácter práctico sobre este tema en nuestro idioma es: Ana Martos Rubio. *Herramientas de búsqueda en Internet*. Madrid: Pearson Education, 2001, 227 p.

Pese a la gran cantidad de publicaciones sobre este aspecto tan importante de Internet, no es frecuente que se dé excesiva relevancia a los metadatos salvo en el caso de publicaciones muy generales sobre el tema, donde su carácter inclusivo remite también someramente al tema de los metadatos. Así ocurre por ejemplo en: Andrew Large, Lucy A. Tedd and R. J. Hartley. *Op. cit.*, donde al tratar el tema de la indización de Internet se refieren a los metadatos, p.136-141; o en: G. G. Chowdhury. *Op. cit.*, uno de los manuales más completos sobre recuperación de información desde la perspectiva bibliotecaria, donde se refiere de soslayo a la metainformación y al Dublin Core en las p. 434-435.

lógicas y semánticas de la información establecidas *a priori* por los creadores de recursos Web.

7.1. Recuperación de información en la Red: consideraciones generales

La RI en Internet es diferente a la que se proyecta y realiza en los servicios de información tradicionales (bibliotecas y archivos, fundamentalmente) ya que la información no se almacena de forma centralizada y porque los recursos de Internet tienen las siguientes características⁴:

- No son fijos y estables como los documentos impresos y otros materiales en soporte físico.
- No están seleccionados y recopilados, como ocurre en las bibliotecas (tradicionales y digitales⁵).
- No están organizados de una forma centralizada para el acceso, ni se catalogan e indizan como los recursos de las bibliotecas.

El mundo de la Web se caracteriza además por ser dinámico y vertiginosamente cambiante, heterogéneo en cuanto a la variedad y la poca uniformidad de la información que alberga, y disperso, porque esa información se encuentra repartida por miles de servidores en distintas partes del mundo y de distintos sectores de actividad; donde cualquiera es o puede ser autor/editor digital. Además de esta veleidad que se le reconoce a la WWW, algunos autores⁶ han

⁴ Kathleen Burnett, Kwong Bor Ng, Soyeon Park. *Op. cit.*, p.1210.

⁵ *Cfr.* Capítulo 8, sobre todo, el apartado 8.2.2.

⁶ Pat Oddy. *Op. cit.*, p. 25.

señalado que su organización es *extremadamente primitiva*. Michael Gorman⁷ por ejemplo, opinaba que Internet *es como una librería inmensa en la que los libros se han apilado de cualquier forma, se han desencuadernado y tienen los índices y las portadas alejadas del resto*.

Al margen de visiones como éstas y de otras más caóticas de algunos "tecnoscépticos", es preciso reconocer que la generalización del uso de Internet, el aumento constante de usuarios, así como la importancia que han cobrado los servicios de información Web, hace que los profesionales de la información, se enfrenten a un mundo informativo pródigo y, en muchos casos, confuso. Los bibliotecarios no pueden ser ajenos a los diversos problemas que plantea la recuperación en un universo de información, en que la sigla WWW que caracteriza este medio, se puede simplificar, cambiando las uves dobles por uves «VVV», teniendo en cuenta las características que, desde nuestro punto de vista, tiene la propia Web: Volumen, Variedad y Volatilidad.

Por otra parte, aunque se manejan muchas cifras, nadie sabe con certeza cuántas páginas conforman la World Wide Web, ni mucho menos cuántas de ellas están visibles o se pueden recuperar a través de los SRII. Uno de los estudios más citados en el ámbito de la cobertura y la búsqueda en la WWW, el de Lawrence y Giles⁸, afirma que en febrero de 1999 se podía acceder públicamente, a través de

⁷[...] *is like a huge used bookstore in which all the books have been piled higgledy-piggledy having been wretched from their bindings and having their indexes and front matter removed*. Citado en: Pat Oddy, *Ibid*. Esta misma idea de Internet como conjunto de libros desordenados, es una metáfora manida en la bibliografía sobre este tema. Aparecía también, por ejemplo, en la introducción sobre los metadatos de la información sobre el proyecto IMS. Educom. IMS Meta-data [documento HTML]. En: *IMSProject*. Instructional Management Systems, rev. 7 de junio de 1998 Disponible en: <http://www.imsproject.org/metadata> (consultado el 17 de julio de 1998).

⁸ Steve Lawrence, Giles, C. Lee. Searching the World Wide Web. *Science*, 3 April 1998, vol. 280, nº 5360, p. 98-100. También disponible en Internet en: <http://www.neci.nj.nec.com/~lawrence/science98.html>. Steve Lawrence, C. Lee Giles. Accessibility of Information on the Web. *Nature*, 8 July 1999, vol. 400, p. 107. Un resumen de los datos de este estudio de Lawrence y Giles se puede consultar en: <http://www.wwwmetrics.com> (consultados el 5 de agosto de 2000).

motores de búsqueda, a unos ochocientos millones de documentos, aproximadamente el 16% de los recursos estimados de la Web. Sin embargo, el trabajo de Dahn⁹ analiza más a fondo estos datos y dice que los resultados de Lawrence y Giles son incorrectos, afirmando que en noviembre de 1999 las páginas Web a las que se podía acceder a través de sistemas de indización de la Web alcanzaban 1,16 billones, mientras que el total estimado de páginas públicamente accesibles se cifraban entre 1,45 y 2,33 billones¹⁰. Dahn añade además que *la Web es una explosión en progreso*, y que *intentar estimar el número de páginas en la Web es como intentar adivinar el número de moléculas arrojadas por un volcán en violenta erupción*. En cualquier caso, normalmente se considera que estos sistemas recogen en sus bases de datos, aproximadamente la mitad de los recursos Web susceptibles de indizarse.

Internet es un medio creciente de diseminación de información que combina muchas de las tareas tradicionales realizadas en bibliotecas y archivos, sin embargo su arquitectura es diferente a la de las unidades de información. Las bibliotecas han sido y son, en primera instancia, sistemas de almacenamiento y recuperación de información y en segundo lugar, un medio de comunicación. A la inversa, Internet es un medio de comunicación y sólo de forma secundaria es un sistema de almacenamiento y recuperación de información. Esta visión de Internet queda patente en la importancia que están cobrando los portales horizontales en la Web, donde la RI *latu sensu* es una parte mínima de los servicios de información que ofrecen.

⁹ Michael Dahn. Counting Angels on a Pinhead: Critically Interpreting Web Size Estimates. *Online*, Jan./Feb. 2000, p. 35-40. También disponible en Internet en: <http://www.onlineinc.com/onlinemag/OL2000/dahn1.html> (consultado el 5 de agosto de 2000).

¹⁰ Un estudio de octubre de 2001, revela que existen en Internet 8,4 millones de *sitios* Web, 1,3 millones más que en el año 2000. On the Size of the World Wide Web [documento HTML]. *Pandia*, 14 de octubre de 2001. Disponible en: <http://www.pandia.com/sw-2001/57-Websize.html> (consultado el 6 de noviembre de 2001). Por su parte, el nuevo buscador Wisenut <<http://www.wisenut.com>> que contabiliza las *páginas* que indiza, tenía registradas 1,571,413,207 páginas en noviembre de 2001. Ambas cifras nos dan una idea del número de sitios y de páginas que puede haber en la Web; sin embargo, estos datos no son tan fiables y reflejan una realidad estrictamente temporal de la Web ya que varían cada día, cada minuto.

A pesar de la inminente fusión de las bibliotecas con el entramado tecnológico e informativo de Internet, todas las características de la Red como fuente de datos, desbordan en cierta manera a los bibliotecarios y documentalistas en cuanto a la organización del conocimiento y recuperación tradicional del mismo (sistemas centralizados basados en un sistema *input-output*, formatos normalizados y persistentes de representación de la información, acceso por materias, existencia de índices estructurados, etc.). Esta desestructuración y falta de centralización de las informaciones difundidas en Internet, ha tratado de ser paliadas, casi desde sus inicios, con herramientas que permitieran ubicar e indizar, de forma masiva, información en la Red. Estas herramientas para la recuperación en Internet han ido desde los propios comandos de unix en los años 80, hasta la creación de agentes inteligentes en la actualidad, siempre manteniendo la tónica de la superación tecnológica, que es la máxima que impera desde que el hombre aplica la informática a la gestión de la información.

Con la popularización de World-Wide Web, y la casi-monopolización de Internet llevada a cabo por este servicio, han surgido diversos tipos de herramientas para la recuperación de información, conocidas esencialmente como directorios y motores de búsqueda, cuyos paradigmas de búsqueda de información son respectivamente, el *browsing* dirigido o navegación, y en la interrogación por palabras clave¹¹. Ambas técnicas o paradigmas de recuperación responden, con frecuencia, a distintas necesidades de información: conocer qué recursos existen en Internet sobre un tema concreto, o bien buscar información puntual sobre un concepto clave. Sin embargo ambos sistemas de búsqueda son susceptibles de la sobrecarga de información y de diferencias del vocabulario, además de otros muchos

¹¹ Existen sistemas de búsqueda y recuperación de información que combinan ambos tipos de búsqueda, la navegación y la interrogación. No obstante, la mayoría de los autores reconocen estos dos paradigmas de forma distintiva y ya tradicional, para caracterizar la recuperación de información en sistemas hipertextuales en red, *v. gr. Browsing hypertext links and performing term-based searching come together as the predominant forms of retrieving information on the Internet*. David Ellis and Nigel Ford. In *Search of the Unknown User: Indexing, Hypertext and the World Wide Web*. *Journal of Documentation*, January 1998, vol. 54, n° 1, p. 41.

problemas detectados en los sistemas de búsqueda general y masiva en la Red, como la fluctuación de los resultados¹² o la cobertura de sus bases de datos.

Todos estos problemas son ya un tópico en la bibliografía sobre RI en Internet que han impulsado el desarrollo de alternativas para mejorar la calidad y fiabilidad de los resultados. Quizás por ello también, han surgido sistemas alternativos y complementarios a los grandes motores y directorios de búsqueda que, basados en metadatos y/o en la estructuración de la información, pretenden paliar estos problemas. En cierta medida muchas de las bibliotecas digitales, en especial las denominadas *subject gateways*, que trataremos en el capítulo siguiente han nacido de la "desconfianza" en estos SRII¹³. También de esa desconfianza o de esa inestabilidad e inconsistencia de la recuperación ha surgido el concepto de Web Semántica¹⁴, así Berners Lee opina que *los aparatos de búsqueda han demostrado ser muy útiles para combinar largos índices rápidamente y para encontrar oscuros*

¹² Mettrop ha estudiado profundamente el tema de la fluctuación de los resultados relevantes en los motores de búsqueda (Altavista, Excite, EuroFerret, HotBot, InfoSeek, Lycos, MSN, NothernLight, Snap, WebCrawler y tres buscadores holandeses: Ilse, Search.nl y Vindex). Argumenta que estas fluctuaciones se deben básicamente a: la existencia de varias bases de datos, de índices diferentes, a que los motores utilizan algoritmos dependientes de las características de los documentos y a que algunos de ellos intentan borrar duplicados de documentos detectados. Wouter Mettrop. Internet Search Engines: Fluctuations in Document Accessibility. *Journal of Documentation*, September 2001, vol. 57, nº 5, p. 623-651.

¹³ La idea de que las *subject gateways* surgen de la inconsistencia de los buscadores de la Red (esto es, SRII) es fácilmente defendible así por ejemplo la encontramos en varios autores como: Phil Bradley. Virtual Libraries and Internet Searching. *Online & CD-ROM Review*, 1999, vol. 23, nº 6, p. 354. Bradley argumenta que, a pesar de la utilidad de los sistemas de búsqueda en Internet, la utilización exclusiva de éstos para responder a una pregunta, es como utilizar sólo una fuente de referencia para atender a una consulta en un centro de información. Estos planteamientos tan elementales de Bradley puede ser una justificación para la utilización de sistemas de recuperación de información especializados, organizados y basados en metadatos desde la perspectiva de la búsqueda (del usuario) que él recoge, pero también pueden entenderse como una motivación para la creación de bibliotecas digitales desde la perspectiva de la creación de estos servicios (bibliotecarios-metadatos-Internet).

¹⁴ Cfr. Capítulo 5, nota 3.

documentos, pero han demostrado ser notablemente inútiles también porque no tienen modo de evaluar la calidad de un documento¹⁵.

A continuación vamos a centrarnos en dos de los aspectos básicos sobre los que se funda la búsqueda de información en la Red: por un lado, la navegación como principio fundamental de búsqueda en los directorios, donde apuntamos una digresión esclarecedora de los distintos tipos de *browsing*; y por otra parte, plantaremos una reflexión sobre el aspecto puntual de la relevancia como problema que subyace a la recuperación de información a través de la interrogación por palabras clave en los motores de búsqueda.

7.1.1. *Browsing* y navegación: aspectos distintivos

En entornos automatizados, pensemos por ejemplo en las bases de datos bibliográficas o en los tradicionales IRS (*Information Retrieval Systems*), donde siempre se han contemplado facilidades de interrogación de distintos tipos. Sin embargo, el paradigma del *browsing* es más nuevo y menos ubicuo en sistemas de recuperación como los OPACs, pero está adquiriendo gran popularidad gracias a la World-Wide Web.

El término *browsing*¹⁶, es uno de los más utilizados, y peor definidos en el contexto de Internet. A pesar de que el *browsing* es una técnica de búsqueda en los OPACs de última generación. Cove y Walsh¹⁷ lo definían como *el arte de no saber lo que busca uno mientras lo encuentra*. De forma más general, se define el *browsing* como *una actividad de búsqueda interactiva en la cual la dirección de la búsqueda*

¹⁵ Tim Berners Lee. *Tejiendo...* Op. cit., p. 163. (La puntuación es nuestra).

¹⁶ Cfr. Glosario.

¹⁷ Citado en: Félix de Moya Anegón. *Los sistemas integrados de gestión bibliotecaria: Estructuras de datos y recuperación de información*. Madrid: ANABAD, 1995, p. 191. Aunque quizás una traducción más fiel, que la que hace este autor de las palabras de Cove, es: *el arte de no saber lo que se busca hasta que se encuentra*.

está determinada por el usuario sobre la base de una inmediata realimentación (feedback) del sistema sobre el que se navega¹⁸. En el entorno de la recuperación Web el concepto de *browsing* tiene una mayor cobertura y distintos matices que debemos aclarar o delimitar. La organización hipertextual de la información implica un tipo de *browsing* especial que se denomina generalmente navegación y para la que nos parece significativa la denominación de "*browsing de valor añadido*"¹⁹ en el sentido de que la organización implícita que subyace a los sistemas basados en él, responde a una aprehensión de metainformación en sentido amplio y sirve de paradigma para la búsqueda de información en directorios o índices.

De forma sistemática, podemos distinguir dos tipos fundamentales de *browsing* en Internet: *surfing* o *browsing* hipertextual y navegación o *browsing* de valor añadido:

- a) El *browsing* hipertextual supone seguir enlaces de forma no sistemática. Siguiendo la perenne metáfora de Internet con el mundo del libro impreso, el *browsing* hipertextual sería como hojear las páginas de un libro, o recorrer los estantes de una biblioteca para obtener un libro de interés.

Siguiendo a Chen y otros autores.²⁰, se puede afirmar que el problema del *browsing* hipertextual surge cuando el modelo mental del usuario no coincide con el modelo establecido en el entramado hipertextual de un conjunto de objetos de información Web de tal forma que el usuario puede experimentar, lo que en la

¹⁸ K. Cox. Information Retrieval by Browsing. En: *International Conference on New Information Technology (5. 1992. Hong Kong)*. Ching-Chih Chen, ed. West Newton: MicroUse Information, 1992, p. 69-79. Citado en: G. G. Chowdhury. *Op. cit.* p. 302.

¹⁹ Ray Denenberg. *Op. cit.*, <http://lcWeb.loc.gov/z3950/agency/papers/italy.html>

²⁰ Hsinchun Chen, et al. Internet Browsing and Searching: User Evaluations of Category Map and Concept Space Techniques. *Journal of the American Society of Information Science*, May 1998, vol. 49, n° 7, p. 584.

terminología de Foss²¹, se conoce como *fenómeno museo de arte* o simplemente, efecto-museo.

Dentro de este tipo podemos identificar, adaptando la tipología tradicional del *browsing* establecida por Cove y Walsh²², dos subtipos o matices:

- *Serendipitous browsing, puramente al azar*. Se refiere a una navegación en la que los usuarios se guían únicamente por el atractivo de los ítems o de los enlaces. En esta línea de la serendipidad²³ o del hallazgo inesperado de recursos relevantes, existen incluso servicios en la Web que basan su localización de información en el mero azar²⁴.
- *Search browsing*: búsqueda directa, donde se conoce el objetivo. Por ejemplo, llegar a un enlace conocido partiendo de una *homepage* inicial en la que el usuario intuye que está el recurso que desea localizar. Para el éxito

²¹ C. L. Foss. Tools for Reading and Browsing Hypertext. *Information Processing and Management*, 1989, nº 25, p. 407-418. Citado en: Chen, et al., *Ibid*.

²² J. F. Cove and B. C. Walsh. Online Text Retrieval via Browsing. *Information Processing and Management*, 1988, vol. 24, nº 1, p. 31-37. Citado en: Lara. D. Catledge. and, James. E. Pitkow. Characterizing Browsing Strategies in the World-Wide Web [documento HTML]. En: *International World-Wide Web Conference (3. 1995. Darmstadt)*. Computer Graphics in Darmstadt: Fraunhofer-Institut für Graphische Datenverarbeitung, abril de 1995. Disponible en: <http://www.igd.fhg.de/www/www95/proceedings/papers/80/userpatterns/UserPatterns.Paper4.formatted.html> (consultado el 17 de julio de 1999).

²³ *Serendipity* es un concepto utilizado en la bibliografía sobre recuperación de información que, desde nuestro punto de vista, resulta muy atinado para denominar aquellos hallazgos informativos que aparecen en los procesos de recuperación de información en Internet por casualidad. Nos ha resultado especialmente interesante el análisis del concepto de "serendipidad" o de recuperación de información por azar en el contexto de las bibliotecas digitales realizado por Eleaine G. Toms. Serendipitous Information Retrieval [documento PDF]. En: *DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries (2000. Zurich)*. ERCIM Workshop Proceedings, Nº 01/W001 [2001]. Disponible en: http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf (consultado el 18 de febrero de 2001). Cfr. Glosario.

²⁴ A este tipo de servicios pertenece, por ejemplo: *Autopilot* <<http://www.mit.edu:8001/people/mkgray/autopilot.html>> constituido por Matthew K Gray, investigador del MIT. Presenta, de manera completamente aleatoria cada 12 segundos, o en el tiempo especificado por el usuario, una página Web de una lista de 8000 páginas seleccionadas cuidadosamente.

en este tipo de búsqueda es más importante el conocimiento de la fuente electrónica, que la propia navegación o que el entramado hipertextual de organización de la misma.

- b) La navegación es el tipo de *browsing* específico y la filosofía de RI que fundamenta índices y directorios de Internet. Denenberg²⁵ se refiere a esta forma de búsqueda como *browsing de valor añadido o browsing con herramientas*. También podemos denominarlo *browsing* dirigido, por contraposición al *browsing* desestructurado y no sistemático que comentamos anteriormente. La navegación, correspondería en el mundo impreso, a la consulta de un catálogo para obtener un libro de una materia deseada o a la consulta de un libro siguiendo el índice.

En la recuperación en el espacio de información Web, esta técnica corresponde a los índices o directorios temáticos, listas donde los recursos aparecen clasificados por materias y submaterias de forma jerárquica (p. ej. Yahoo <<http://www.yahoo.com>>).

Chen y otros²⁶, ven en este tipo *browsing*, que denominan de navegación-directorio (*directory-browsing*), un método para mejorar la eficacia en la búsqueda por el amplio espacio de información de la Red. La categorización o la clasificación que se sigue en estos directorios es similar a la tendencia de organizar el conocimiento en las bibliotecas mediante sistemas de clasificación. Sin embargo, las categorías de estos sistemas son limitadas en cuanto al nivel de

²⁵ *Browsing refers to the unstructured, serendipitous process whereby a user, hoping to arrive at a document of interest, manoeuvres from link-to-link (or page-to-page) without any informed tools to help decide what link to follow. Navigation is perhaps best described as value-added browsing, or browsing with tools.* Ray Denenberg. *Op. cit.*, <http://lcweb.loc.gov/z3950/agency/papers/italy.html>

²⁶ Hsinchun Chen, et al., *Loc. cit.*

especificidad y oportunidad. Por ello, para mejorar las clasificaciones que se establecen se tiende a utilizar la clasificación automática utilizando algoritmos²⁷.

Volviendo a la tipología tradicional de Cove y Walsh²⁸ para completar esta visión de la navegación como paradigma de búsqueda en los índices temáticos, ese *browsing-directorio*, *browsing de valor añadido* o navegación, corresponde a lo que ellos denominan *browsing de propósito general* (*general purpose browsing*) que implica la consulta de fuentes que tienen una alta probabilidad de contener ítems de interés.

En cualquier caso, comentamos aquí esta técnica como sistema de recuperación de información porque intuitivamente, la navegación o *browsing-directorio* y la interrogación no son actividades que se excluyan mutuamente, incluso además, cualquier estrategia de búsqueda está constantemente influida por el *browsing*²⁹. Asimismo, la navegación es la forma de recuperación de información que subyace a algunas de las nuevas tendencias para la organización y recuperación de recursos electrónicos en red como los mapas de materia (*Topic Maps*) o las

²⁷ Hsinchun Chen, et al., *Ibid.* Estos autores, hablan de un algoritmo concreto desarrollado con esta finalidad, denominado *Kohonen self-organizing feature map*. Al analizar las tendencias en los sistemas de recuperación de información en Internet (SRII) desde una perspectiva semántica, hablaremos de algunos sistemas que tratan de emular más estrictamente los sistemas de organización bibliotecaria y establecen índices temáticos. *Cfr.* 7.2.

²⁸ Citado en: Lara. D. Catledge and James. E. Pitkow. *Op. cit.*, <http://www.igd.fhg.de/www/www95/proceedings/papers/80/userpatterns/UserPatterns.Paper4.formatted.html>.

²⁹ De hecho, se están diseñando sistemas de recuperación de información que combinan ambas técnicas, tal es el caso de WebGlimpse <<http://webglimpse.org>>, uno de los software gratuitos parte del proyecto Harvest, basado en la conjunción de ambos paradigmas (*browsing* y *searching*), originalmente desarrollado la Universidad de Arizona. Udi Manber, Burra Gopal, and Michael Smith. Univ. de Arizona Position Paper: Combining Browsing and Searching. En: *Distributed Indexing/Searching Workshop (1996. Cambridge)*. World Wide Web Consortium, 2 de junio de 1996, rev. 20 de junio de 1996 [documento RTF]. Michael Schwartz, Mic Bowman, dir. World Wide Web Consortium, 1996, rev. 18 de julio de 1996. Disponible en: <http://www.w3.org/Search/9605-Indexing-Workshop/Papers/AllPaps.doc>, p. 71. Este sistema se basa en la introducción del concepto de *neighborhoods* —entendidos como contextos de información próximos a los de una página— y en el *browsing* en estos "dominios próximos" posteriormente a la recuperación de información concreta.

ontologías. Los mapas de materia se utilizan *para organizar la información de tal forma que se pueda optimizar la navegación*³⁰. Ambas formas de organización informática del conocimiento se han diseñado —como ocurre con los metadatos orientados al recurso— para resolver el problema de la gran cantidad de información no organizada, pero en este caso, desde una perspectiva de metadatos orientados a la materia o *metadatos descriptivos de la semántica*³¹.

7.1.2. Paradigma de interrogación: el problema de la relevancia

En líneas generales, la interrogación se puede caracterizar como un proceso en el que el usuario lanza a una base de datos una sentencia de búsqueda utilizando palabras clave y el sistema debe localizar la información que mejor satisfaga la búsqueda a través del emparejamiento (*matching*) entre el/los términos de la sentencia, y el/los términos almacenados en la base de datos.

El propósito de todo sistema de representación y recuperación de información (independientemente de que sea un catálogo, una base de datos bibliográfica, o la base de datos de un buscador de la Red) no es proporcionar mucha información, sino proporcionar información relevante. Por lo cual, el concepto de relevancia se sitúa en el centro de la discusión de este estudio, teniendo en cuenta que, dada la sobrecarga de información de la Red, cada vez es más difícil obtener recursos relevantes.

Un estudio exhaustivo de los problemas de la recuperación que acontecen en un SRII de este tipo (motores de búsqueda), rebasaría los objetivos de este trabajo, pero

³⁰ ISO13250-1999 *Document Description and Processing Languages - Topic Maps*. Citado en: Eric Freese. *TopicMap vs. RDF* [documento HTML]. London: Open University, 2000. Disponible en: <http://ep.open.ac.uk/PubSys/resources/html/free0000.html> (consultado el 17 de julio de 2001).

³¹ Terje Brasethvik. A Semantic Modeling Approach to Metadata. *Internet Research: Electronic Networking Applications and Policy*, 1998, vol. 8, n° 5, p. 381. Brasethvik denomina a estos modelos de recuperación de información basados en metadatos de contenido, *semantic descriptive metadata*, para referirse a aquellos que directamente evocan el contenido intelectual de un DLO. En sentido amplio, los *topic maps* y las ontologías responden a esta aprehensión semántica y contextual de la recuperación de información.

sí nos parece inevitable hacer una reflexión puntual en torno al problema de la relevancia que se produce en estos sistemas basados en el paradigma de la interrogación. Podemos definir la relevancia, como la relación que existe entre una necesidad de información manifiesta y la información almacenada. Esta definición tan sencilla alberga una gran trascendencia en el contexto de Internet. Por una parte, en torno al concepto de información almacenada, que en este caso no es un almacenamiento "tangible"³² ni unívoco, y por otro, en torno a lo que hemos llamado necesidad de información manifiesta, ya que en Internet, muchas veces no existe tal necesidad concreta de información.

Tradicionalmente en los servicios de información, la relevancia se ha medido a través del número de documentos que satisfacen una necesidad de información de un usuario, y ha dado lugar a indicadores y parámetros de evaluación cuantitativa del éxito de un sistema de recuperación de información a través del cálculo de la tasa exhaustividad:

$$\frac{\text{N}^\circ \text{ de documentos relevantes recuperados por el sistema}}{\text{N}^\circ \text{ total de documentos relevantes contenidos en el sistema}^{33}} \cdot \%$$

Esta ratio no es operativa en Internet dado el crecimiento exponencial de la información y el resto de las particularidades de la información hipertextual de la

³² Aunque Pat Oddy, desde una perspectiva clásica señala que es una falacia creer que los medios electrónicos no tienen existencia física. *Todo documento electrónico está almacenado de alguna forma para un acceso y recuperación posterior [...] la única diferencia es que el sistema de almacenamiento no tiene que ser necesariamente visible.* Pat Oddy. *Op. cit.* Con la afirmación de que el almacenamiento de la información de la Red no es tangible, nos referimos a la diferencia entre el almacenamiento en forma de átomos de los libros y documentos impreso, frente al almacenamiento en bits de los documentos electrónicos, lo que dificulta, en nuestra opinión, una recuperación de información relevante, cuanto menos desde el punto de vista de la no-asibilidad de la información, que dificulta su control.

³³ Frederick W. Lancaster. *Op. cit.*, p. 152.

WWW³⁴. Cuando los resultados que responden a una búsqueda son muy elevados (casi siempre), los motores de búsqueda hacen un ranking de relevancia, ordenando la lista de resultados de tal forma que las páginas que *posiblemente* sean más adecuadas para la pregunta, aparezcan en primer lugar. En cualquier caso, la relevancia de un sistema de recuperación de información nunca es absoluta, y en la Web, al calcularse ésta sobre los documentos indizados en texto completo de los documentos, aún es más relativa dada la polisemia y anfibología del lenguaje natural en el que están los contenidos de la WWW. En el glosario especializado del servicio SearchTools³⁵ se define relevancia como *la medida de cuán bien una página indizada responde a una pregunta*, y matizan además que sólo el usuario (*searcher*) puede definirla realmente.

Para evaluar la relevancia, es preciso pues, tener en cuenta una doble perspectiva: por un lado, el punto de vista cognitivo sobre cómo el opera y se comporta el usuario ante una necesidad de información; y por otro, el punto de vista puramente algorítmico. Pese a que ambas visiones no son excluyentes, no siempre se simultanean en un proceso de interrogación-recuperación. De esta forma, distinguimos entre relevancia subjetiva (la del usuario), y relevancia objetiva o valorable (la del sistema)³⁶:

a) Relevancia subjetiva. Para desarrollar esta idea, partimos de que el concepto de relevancia es subjetivo en sí mismo y de que el usuario siempre tiene una intención concreta (una necesidad manifiesta de información). Cuando un usuario

³⁴ A estas particularidades nos referimos anteriormente con la metáfora visual de "VVV": Volumen, Variedad y Volatilidad.

³⁵ Search Terms Glossary (SearchTools) <http://www.searchtools.com/info/glossary.html#relevance>

³⁶ Algunos autores, partiendo de una base similar a la que expresamos aquí, distinguen entre relevancia semántica (subjetiva, para nosotros) y relevancia formal (objetiva). *Vid.* Alfons Cornellá. La importancia de la relevancia en información. *Extra-Net! / IWETEL* [correo electrónico en lista de distribución]. 20 de enero de 1998, 00:18:00. Disponible en: <http://www.extra-net.net/articulos/en980119.htm> (consultado el 2 de agosto de 1998).

realiza una consulta específica, él es intencionadamente relevante, porque sabe lo que está pensando y lo que quiere encontrar y espera lo mismo del sistema de búsqueda que interroga. Pese a ello, hay que tener en cuenta que el usuario no puede modificar de un modo efectivo el contexto o entorno cognitivo de su interlocutor —el motor de búsqueda— que es, en la mayoría de los sistemas más populares de búsqueda global en Internet = 0. No puede transmitir al buscador los estímulos ostensivos que evidenciarían su intención determinada³⁷. Y por ello, del *ranking* de resultados que el sistema le devuelve (según la relevancia objetiva —*Cfr. infr.*—) no tienen por qué ser los diez primeros, los más pertinentes para él (relevancia subjetiva), y sí quizás, encontrar lo que buscaba en el último enlace de la lista de resultados que presenta el sistema.

Cuando el motor de búsqueda devuelve un conjunto de ítems o enlaces, todos son relevantes para la palabra/s-clave propuestas en la consulta, pero es el usuario el encargado de medir hasta qué punto esas respuestas son relevantes para sus expectativas, es decir, determinar qué precisión ha tenido la respuesta del buscador³⁸.

- b) Relevancia objetiva. Esta perspectiva de la relevancia se basa en el análisis del propio mecanismo de recuperación, esto es, del algoritmo/s que utiliza el *search engine*³⁹, que es transparente para el usuario.

³⁷ Esta reflexión se basa en la teoría de la relevancia de Sperber y Wilson que tiene un enfoque cognitivo. *Cfr.* Sebastián Bonilla Álvarez. Información y relevancia. Hipótesis acerca de cómo procesamos los seres humanos la información lingüística. En: *Revista Española de Documentación Científica*, 1996, vol. 19, nº 4, p. 392-410.

³⁸ La precisión es otro de los conceptos básicos en la evaluación de un SRI y se define como la ratio del número de ítems relevantes encontrados fruto de una búsqueda documental, sobre el total de los ítems encontrados:

$$\frac{\text{Nº de documentos relevantes recuperados}}{\text{Nº total de docs. recuperados}} \%$$

³⁹ Para una explicación exhaustiva de este anglicismo y de su uso, *Vid.* Glosario.

Desde este punto de vista, lo que se tendrá que evaluar es el sistema/software que realiza el cálculo de la relevancia, que puede ir desde el n° de veces de los términos de la pregunta que aparecen en cada documento candidato para satisfacerla (*exact matching*), al cálculo más sofisticado utilizando algoritmos de asignación de pesos a los términos⁴⁰ que ponderan con un valor más alto a los más específicos (menor frecuencia global), y con un valor más bajo a los muy abundantes. También hay motores que para la recuperación se basan en el método de asignación de un valor vectorial a los términos de la pregunta, y calculan el coeficiente de similitud comparándolo con los vectores de cada término del índice invertido⁴¹.

Para esos métodos del análisis de la relevancia (asignación de pesos o del cálculo de un valor vectorial), los *search engines*, pueden utilizar metadatos o *pseudo-metadatos* embebidos en la propia sintaxis HTML del documento, que permiten otorgar mayores pesos o mayores valores a los términos definidos en las etiquetas <META>, donde se ubica, como hemos visto, dentro de la cabecera del DLO y en forma de pares de atributo-valor, la metainformación.

Ante esta doble caracterización de la relevancia, parece obvio que la asignación de metadatos a los objetos de la Red, permitirá una mayor relevancia en los sistemas de recuperación global, tanto desde un punto de vista objetivo como subjetivo. Esos metadatos de carácter general servirán, teniendo en cuenta que en lo que se fija el usuario⁴², para determinar la utilidad de los recursos recuperados como resultado a

⁴⁰ Sobre este aspecto, *Vid. Information Retrieval: Data Structures & Algorithms*. William B Frakes, Ricardo Baeza-Yates, eds. New Jersey: Prentice Hall, 1992, p. 371.

⁴¹ El cálculo de la relevancia se realizará así, cuando el modelo de indización seguido por el *search engine* sea el modelo del valor de discriminación o modelo de espacio vectorial, como es el caso de WebCrawler <<http://webcrawler.com>>, que permite ponderar los términos en el momento de la indización, según su frecuencia de aparición en el texto. Sin embargo otros *search engines*, como el que utiliza Excite <<http://www.excite.com>>, se basan en algoritmos de clusterización o agrupación.

⁴² Lo que valora un usuario de una búsqueda, tiene relación con la apreciación de Dempsey y Heery sobre lo que necesitan saber los usuarios sobre la información que buscan en Internet (*Cfr.* 3.4. *Vid.* [cont.]

una consulta (relevancia subjetiva). De esta forma, la relevancia dependerá no sólo de la función de comparación de la palabra clave introducida, con el conjunto de palabras clave existentes en el índice invertido del buscador en cuestión; dependerá en muchos casos también, del tipo de documento encontrado, del autor, de la actualidad del recurso (fecha), de la lengua en que esté escrito, etc. y de la medida en que el resultado pueda anticipar al usuario el interés que tienen los documentos encontrados potencialmente para él⁴³.

Parece evidente pues que, la consignación estructurada de esos datos sobre el documento —metadatos sobre: autor, título, tipo de DLO, palabras clave, resumen, etc.— codificados en el propio DLO, son esenciales para que un *search engine* de carácter genérico pueda aplicar sus algoritmos de recuperación basándose en ellos, ya que el usuario tiene en cuenta esa información para determinar la relevancia subjetiva de los resultados. No obstante, como explicaremos a continuación, las tendencias de los SRII de carácter global son muy distintas y están en constante evolución, al mismo tiempo que cambia la concepción de los metadatos y la aprehensión que hacen de ellos los usuarios, los gestores de información y los diseñadores de aplicaciones.

Loncan Dempsey and Rachel Heery. Metadata: A Current View... *Op. cit.*, p. 148-149), ya que pueden ser estas mismas necesidades las que sean determinantes a la hora de evaluar la relevancia subjetiva de los resultados. Si en el apartado 3.4., estas necesidad nos mostraban las distintas aplicaciones de los metadatos, ahora nos sirve para argumentar la necesidad de que éstos se establezcan para mejorar la búsqueda global en Internet, ya que el usuario puede determinar así, en qué medida le interesa seguir un enlace de un DLO recuperado y relevante desde el punto de vista objetivo.

⁴³ Al hablar de la medida en que un resultado de una búsqueda pueda anticipar el contenido, y con ello, el interés de un documento para un usuario concreto, nos estamos refiriendo a la información que el sistema devuelve sobre los documentos encontrados, que no es otra que la que previamente se haya recabado de manera automática, o se haya introducido en forma de metadatos. Sobre este aspecto es muy interesante la aportación de: A. Wheatley and C. J. Armstrong. Metadata, Recall, and Abstracts: Can Abstracts Ever be Reliable Indicators of Document Value? *Aslib Proceedings*, September 1997, vol. 49, nº 8, p. 206-213. En este interesante artículo, los autores nos invitan a reflexionar acerca del papel que desempeña el resumen documental (atributo *description*) y la indización (atributo *keywords*) o sus análogos funcionales según los distintos modelos que mencionamos en el capítulo 5, en el contexto de la recuperación de información global en la Red.

7.2. Tendencias de los sistemas de búsqueda Web... ¿metadatos?

Aunque llevar a cabo un análisis minucioso de las herramientas de búsqueda para la Web no es objetivo de esta investigación, sí creemos necesario plantear en líneas generales, la pléyade de sistemas existentes, así como las tendencias en ellos con relación a los metadatos. El hecho de incluir este apartado aquí tiene una doble función: de un lado, presentar el panorama de la búsqueda y localización de información Web, de tal forma que se justifique la necesidad de una Web más semántica basada en estructuras de metadatos; y de otro, plantear un contrapunto de RI a las bibliotecas digitales que trataremos en el capítulo siguiente.

Las herramientas de búsqueda y recuperación de información en la Red (SRII⁴⁴) se fundamentan en la automatización de las tareas tradicionales de organización del conocimiento (clasificación e indización). De forma elemental, podemos decir que un software clasifica y/o indiza automáticamente los datos de Internet, organizando de alguna manera la miríada de información digital existente.

El tratamiento automático a la información, evita el alto coste y los retrasos y la imposibilidad de realización, que supondría una indización humana de todos los recursos distribuidos en la Web. Sin embargo y a diferencia de los indizadores humanos, estas aplicaciones, en general, no identifican características de un documento como la materia de la que trata, el autor, la fecha de publicación, el tipo de documento o las condiciones de acceso (no pueden distinguir por ejemplo, si un documento es un poema o un informe científico, etc.). Todo esto se debe a que la

⁴⁴ Es preciso advertir que dentro de lo que denominamos aquí SRII, sólo contemplamos aquellos sistemas destinados a la recuperación de recursos Web en general, no incluimos otros sistemas de recuperación no-Web que también tienen acceso a través de la misma. Tratamos los servicios de búsqueda tal y como se conocen popularmente, aunque al hablar de motores de búsqueda, mencionaremos que el *search engine* propiamente dicho, se puede aplicar a información electrónica local, independientemente de que esos *engines* sean, en muchos casos, la base de un servicio de recuperación global. Cfr. Glosario, *search engine*. Esta metodología nos permitirá confrontar sistemas de recuperación global (capítulo 7) y sistemas de recuperación local, contextos informativos definidos y finitos, o bibliotecas digitales (capítulo 8).

mayor parte de los *engines* que indizan la Red, no registran metadatos; y en el caso de que sí estén preparados para reconocerlos, los autores/editores de las páginas Web no utilizan un modelo de metadatos de forma generalizada, que permita insertar en la cabecera del documento esta información, determinante para el usuario, tanto a la hora de realizar sus consultas, como a la hora de comprobar si los resultados son relevantes según sus expectativas, sin necesidad de seguir el enlace y acceder al documento completo.

Parece oportuno evocar también aquí la afirmación de Jones en los años 80: [...] *el valor de la indización automática se incrementará cuando la literatura de forma legible a máquina sea más importante que la producida por medios tradicionales [...]*⁴⁵. La información legible por máquina no es más importante que la bibliografía impresa, pero es obvio que juega un papel esencial en la nueva concepción de sistemas de información. Esta situación aumentará indudablemente en los próximos años, ya que cada vez más existen proyectos de digitalización de materiales impresos, múltiples publicaciones se realizan en ambos formatos (impreso y digital) y cada vez más también, se crean documentos originalmente en formato electrónico que no tienen un homólogo impreso.

A continuación vamos a plantear un acercamiento a todas esas formas y tendencias de recuperación de información en la Red. Para ello, proponemos una doble aproximación⁴⁶ a los SRII:

⁴⁵ K. P. Jones. Getting Started in Computerized Indexing. *The Indexer*, 1986, vol. 15, nº 1, p. 12. Citado en: José Antonio Moreiro González, Eva M^a Méndez Rodríguez. *Op. cit.*, p. 14.

⁴⁶ Una aproximación similar a la que aquí señalamos es la que defienden Corrado Pettenati y Carla Basili en el capítulo 5 de su obra *La biblioteca virtuale*. Citado en: Gabriele Lunati. Strumenti per la ricerca ed il recupero dell'informatizzazione su Internet. *Workshop Electronic Information in Libraries and Documentation Centres (5. 1996. Milán)*. En: Ricardo Ridi. *ESB Forum: Recensioni e contributi su cd-rom e altre fonti informative elettroniche*. Andrea Castagna e Patrizia Rivara (col.). Genova: E.S. Burioni Ricerche Bibliografiche, 1996. También disponible en Internet en: <http://www.burioni.it/forum/motori/par8.htm>

- Por un lado, una aproximación práctico-tecnológica que se realiza gracias a una estrategia que usa la potencia de cálculo y la capacidad de transmisión de las redes para la recuperación de un volumen ingente de datos e información, lo que los investigadores de la *Alexandria Digital Library* denominan *método de fuerza bruta*⁴⁷.
- Por otro, una aproximación teórico-semántica, de cariz bibliotecario, según la cual el acceso a los recursos mejora imponiendo a los datos una estructura. En este sentido analizaremos lo que hemos denominado sistemas de búsqueda global "de organización bibliotecaria"⁴⁸.

Básica y tradicionalmente ya, con el afán de simplificar la compleja e inabarcable realidad de la recuperación de información en la WWW, se distinguen los siguientes tipos de SRII⁴⁹: Directorios o índices, motores de búsqueda y metabuscadores, junto a la nueva y popular realidad de los portales que contemplan

⁴⁷ *The brute-force method of searching the Internet, such as Altavista, gathers and indexes "all" documents (i.e. information objects) on the network.* Linda L. Hill, et al. Collection Metadata... *Op. cit.*, p. 1177.

⁴⁸ Esta es la aproximación que se trata de refrendar y matizar a lo largo de toda investigación, ya que la estructuración de la información a través de metadatos es una tendencia en esa línea teórico-semántica.

⁴⁹ Para una caracterización más exhaustiva y certera sería preciso ampliar esta investigación al caso individual de cada sistema de recuperación de información de Internet, pormenorizando las características de: software que utilizan para la indización automática, la intervención o no del hombre en esa indización, la cobertura, el proceso de búsqueda, el cálculo de la relevancia, etc. Existen varios estudios en esta línea. Resulta muy interesante por ejemplo, el trabajo realizado por Traugott Koch. Detailed Description of Important Search [documento HTML]. En: *Searching the Web. Systematic Overview Over Indexes Services*. Lund University Electronic Library, febrero de 1995, rev. 25 de febrero de 1998. Disponible en: http://www.ub2.lu.se/tk/websearch_systemat.html#detai (consultado el 2 de julio de 2001). Este recurso, a pesar de ser riguroso y exhaustivo desde la perspectiva de la documentación, no se actualiza desde 1998. Para una información actualizada y certera sobre el estado de los sistemas de búsqueda recuperación de información en Internet, es fundamental la consulta de:

- Searchengineshowdown: <http://www.searchengineshowdown.com>
- Searchenginewatch: <http://www.searchenginewatch.com>
- SearchTools: <http://www.searchtools.com>

la búsqueda de información como un servicio más que se funda en directorios, en motores de búsqueda, o en ambos. Realizaremos una valoración de los mismos desde la perspectiva de los metadatos, en aras a reflejar una implementación realista de la metainformación en la búsqueda genérica o global de la Web.

En primer lugar, los directorios ó índices temáticos, son grandes repertorios jerárquicos que se basan en la clasificación humana de los distintos recursos de Internet, y presentan como resultado unos "catálogos" o listas temáticas. Se trata de sistemas pre-coordinados de recuperación de información, ya que ofrecen inventarios de materias que parecen, en muchos casos, trasuntos de las clasificaciones bibliotecarias tradicionales⁵⁰ o listas pre-coordinadas de encabezamientos de materia. Como modelo paradigmático de estos sistemas, se ha tenido siempre a Yahoo! <<http://www.yahoo.com>> que, a pesar de haberse "portalizado" (es decir, de haberse convertido en un servicio de valor añadido y una puerta genérica de acceso a Internet) mantiene sus catorce categorías originales, donde trata de organizar, con ayuda de otras subcategorías, "todos" los recursos de la Web.

La técnica de búsqueda y recuperación de información en estos sistemas, combina, en muchos casos, la técnica de la navegación dirigida (*general purpose browsing*) con la búsqueda automatizada por palabras clave. Sistemas como Yahoo, que utilizamos como prototipo de estos SRII, pueden responder a una búsqueda post-coordinada por palabra clave, a través de la presentación de cadenas de términos de indización pre-coordinados que representan una clasificación de materias complejas, ordenadas jerárquicamente, donde se puede además reducir la búsqueda por término de indización a una categoría determinada, lo que de alguna forma no deja de ser una

⁵⁰ Existen algunos autores que, tratando de ver en estos índices temáticos como trasuntos de las clasificaciones bibliotecarias, han establecido comparaciones tan peculiares como la establecida por Glassel entre la Clasificación Colonada de Ranganathan y el directorio Yahoo. Aimee Glassel. Was Ranganathan a Yahoo? [documento HTML]. *InterNIC News*, 1998, nº 3. Disponible en: <http://scout.cs.wisc.edu/addserv/toolkit/enduser/archive/1998/euc-9803.html> (consultado el 28 de noviembre de 2001).

restricción semántica de la búsqueda⁵¹. Estos índices temáticos suelen ser una herramienta útil para iniciar una búsqueda, pero su validación humana y la subjetividad que esto comporta no le permiten asumir el crecimiento incesante de la Red.

No podemos hablar en sentido estricto de metadatos asociados a este tipo de SRII, sin embargo desde la perspectiva conceptual del "todo metadatos" que señalamos en el capítulo 3⁵², las categorías clasificatorias que establecen estos sistemas, podrían entenderse como metainformación que anticipa, en cierta manera, el contenido de los recursos clasificados dentro de cada tema, al agrupar la información de "en listas pre-coordinadas de encabezamientos de materia". La organización de los recursos que realizan estos directorios nos podría llevar a considerarlos sistemas de recuperación de información "de organización bibliotecaria"; aunque no son bibliotecas digitales⁵³ en sentido estricto, sí se basan en criterios semánticos de organización del conocimiento y en muchos casos, surgen de instancias expresamente bibliotecarias. Por ejemplo Digital Librarian <<http://www.digital-librarian.com>>, que explícitamente plantea "una elección hecha por bibliotecarios a lo mejor de la Web" con más de ochenta categorías interrelacionadas donde se recogen recursos seleccionados y a los que se puede acceder a través de la navegación o *browsing* de valor añadido.

⁵¹ Aunque la creación de metadatos está abocada a paliar estos dos problemas, a través de estructuras de información, las soluciones más contundentes en este sentido están dirigidas al desarrollo de potentes algoritmos. Un ejemplo del desarrollo algorítmico para solucionar los problemas del *browsing* (*algoritmo de mapa de categorías*) y de la interrogación (*algoritmo de espacio conceptual*) se puede ver en: Hsinchun Chen, et al. *Op. cit.*

⁵² *Cfr.* 3.2.

⁵³ *Cfr.* Capítulo 8.

La categorización temática puede ser resultado de los metadatos como vimos en los *pathfinders* de CORC o en las utilidades de algunas de las aplicaciones para la gestión de metadatos⁵⁴. Pero sólo podemos hablar de metadatos *strictu sensu*, vinculados a directorios generales en el caso de Open Directory Project (ODP) <<http://dmoz.org>> un directorio "puro" (en el sentido de que no se ha portalizado). Creado en junio de 1998 e impulsado por Tim Bray y Netscape-Mozilla, se basa en la misma idea que Yahoo de clasificar la información, pero aplicando métodos de fuente abierta (similares a la filosofía del software *open source*, e incluso mucho más democrático pues no requiere ningún tipo de programación) que sirven para la clasificación cooperativa de información utilizando RDF.

Según Hjelm⁵⁵ ODP pretende ser *una república autorregulada donde los expertos pueden recopilar sus recomendaciones, sin incluir ruido, ni información errónea*. Permite descargar libremente (previa licencia) sus conjuntos de datos (*dumps*) en RDF⁵⁶, utilizando las etiquetas del propio *schema* de ODP⁵⁷. El objetivo es crear un directorio global de la Web, confiado a editores voluntarios y al espíritu de colaboración inmanente a Internet. Con más de tres millones de sitios en su base de datos, con una organización basada en metadatos RDF generados *a posteriori* de la creación del recurso y que sustentan la estructura de directorio, Open Directory constituye, desde nuestro punto de vista, una primera y tímida visión de la Web Semántica, ya que su cobertura es general e implica la organización de los recursos de la WWW, y permite resultados consecuentes en las búsquedas. La existencia de proyectos como éste, no invalida nuestra teoría de aplicabilidad de metadatos a

⁵⁴ Cfr. Capítulo 4, nota 171. Vid. Glosario, *pathfinders*.

⁵⁵ Johan Hjelm. *Op. cit.*, p. 134.

⁵⁶ Cfr. *infr.* Oingo <<http://www.oingo.com>>. Los datos que se pueden descargar (*dumps*) de ODP, no son totalmente XML válido pues este proyecto se inició cuando todavía la especificación de RDFMS era un borrador. Cfr. <http://dmoz.org/rdf/Changes.html> y <http://www.w3.org/RDF>

⁵⁷ *RDF Tags Used in the Current RDF Files*: <http://dmoz.org/rdf/tags.html> (Consultados en la fecha de su actualización 27-11-2001).

bibliotecas digitales, sino que anticipa una proyección de los metadatos a la búsqueda Web global.

El otro tipo de SRII que funcionan de forma global sobre la Web, los motores de búsqueda o *search engines*⁵⁸, son sistemas post-coordinados de recuperación de información en Internet, que se basan en la búsqueda por palabras clave. Realizan una indización sobre el texto completo o parte del texto de los documentos Web basándose fundamentalmente en técnicas de análisis estadístico y/o ponderación y presentan los resultados organizados según un ranking de relevancia⁵⁹. Los componentes que integran estos SRII, son fundamentalmente los mismos que para cualquier sistema de recuperación en general⁶⁰: una base de datos constituida a partir de la indización del texto completo de los documentos, un software o robot que recorre automáticamente la Red, detectando nuevos enlaces, un indizador de recursos, un *search engine* y una interfaz de usuario que permite formular y lanzar sus búsquedas y visualizar los resultados. En cuanto al proceso de emparejamiento (*matching*) entre los términos de la consulta del usuario y los recursos indizados en la base de datos del buscador, dependerá de sus respectivos lenguajes de interrogación aunque la mayoría permiten el emparejamiento exacto de palabras clave o frases, la formulación de consultas a través de operadores boléanos y/o relacionales, truncamiento y/o adyacencia.

⁵⁸ Los motores de búsqueda reciben múltiples denominaciones en inglés, muchas de ellas, fruto del lenguaje metafórico que acompaña a la creación semántica sobre Internet. *Vid.* Glosario, *search engine*, *searchbot*.

⁵⁹ Aunque aquí sólo incidiremos en la importancia de los metadatos y/o metaetiquetas para el ranking de relevancia de los motores de búsqueda, Bradley analiza todos los condicionantes que operan en estos sistemas para el cálculo de la relevancia. *Vid.* Phil Bradley. The Relevance of Underpants to Searching the Web [documento HTML]. *Ariadne*, 23-Jun-2000, issue 24. Disponible en: <http://www.ariadne.ac.uk/issue24/search-engines> (consultado el 5 de julio de 2001).

⁶⁰ *Vid.* David Ellis and Nigel Ford. *Op. cit.*, p. 28. *Cfr.* 8.1.

Al comentar la tipología de metadatos de Depsey y Heery⁶¹, así como al hablar de la edición de metaetiquetas⁶², comentamos que estos servicios operan fundamentalmente con los *pseudo-metadatos* que clasificábamos en la banda uno, con lo cual no existe una verdadera estructuración previa de la información, sino que los datos sobre los datos se extraen de manera automática y sin diferenciación. Jian Qin⁶³ interpreta esta extracción automática de información como un tipo de creación de metadatos que él denomina *representación post-publicación* (*Post-publishing representation*), definiéndola como *un método en el cual un tipo especial de programa de ordenador genera metadatos de un objeto digital ya publicado*. Sin embargo, esta afirmación de Qin es falaz, pues se puede interpretar como que los buscadores extraen metainformación automáticamente (lo que podía asimilarse a los modelos de extracción de Klarity o de DC-Dot⁶⁴). En realidad esta visión es sólo una generalización más de los metadatos, que lleva la interpretación *latu sensu* de los mismos ("todo metadatos") a sus más amplias consecuencias, ya que considera la indización automática como una extracción de metadatos.

Los "metadatos" que conforman los "registros" construidos de forma automatizada en estos sistemas de búsqueda, no son lo suficientemente exhaustivos para permitir al usuario hacer juicios previos sobre la relevancia (subjetiva) de los recursos efectivamente recuperados. Normalmente los software rastreadores (robots) de estos sistemas de búsqueda no son selectivos con los recursos que indizan, y por

⁶¹ Lorcan Dempsey and Rachel Heery. *A Review of Metadata...* Op. cit., <http://www.ukoln.ac.uk/metadata/desire/overview.rtf>, p. 8-9. Cfr. 3.5.

⁶² Vid. 4.3.1.

⁶³ Por contraposición a la asignación de metadatos que denomina estructuración prepublicación (*Pre-publishing structuring*). Jian Qin. Representation and Organization of Information in the Web Space. *Informing Science*, 2000, vol. 3, nº 2, p. 84. También disponible en Internet en: <http://inform.nu/Articles/Vol3/v3n2p83-88.pdf>

⁶⁴ Cfr. 4.2. y 4.3.2.1.

todo ello, a pesar de la popularización de su uso para la RI en el entorno de Internet, son sistemas deficientes desde el punto de vista de la eficacia en la recuperación.

Los metadatos que se tienen en cuenta en la indización (y en su caso, en el cálculo de la relevancia), son lo que denominamos "metaetiquetas"⁶⁵, creadas como estructuras genéricas de metainformación (cabeceras HTML), por los autores/editores de los recursos Web. Si estos sistemas utilizan metadatos HTML⁶⁶ para indizar

⁶⁵ Vid. 4.3. y 5.2.1.1. Cfr. Glosario, *metatag*.

⁶⁶ No existe un acuerdo sobre los motores de búsqueda que utilizan las etiquetas <META> para el cálculo de relevancia en sus búsquedas ya que normalmente el funcionamiento interno de los buscadores es transparente para los usuarios y los propios creadores de este tipo de software tampoco hacen excesiva publicidad de sus funcionalidades. Algunos ejemplos sobre este disenso en los motores que tienen en cuenta las etiquetas <META> para la indización son:

- Según Koch, los sistemas de recuperación global que ofrecían (1998) que se basaban, de alguna forma, en metaetiquetas HTML para la indización, eran: Altavista, Infoseek, SwishSearch, Hotbot, Webcrawler, los motores de búsqueda de información en alemán, Aladin y Eule y los SRII especiales para metainformación, Fireball, Polish Academic Web Index, HotMeta Search Engine (que indizaba metadatos DC) y Nordic Metadata Index (que indizaba los modelos IAFA, DC y GILS). Traugott Koch. *Documentation to: Metadata Indexing and Searching in Large Search Services*. [documento HTML]. Lund University Electronic Library, 8 de enero de 1998, rev. 13 de febrero de 1998. Disponible en: <http://www.ub2.lu.se/~traugott/Mdsearch-docu.html> (consultado el 28 de julio de 1998).
- Para Rowland, los motores que soportan metaetiquetas son Altavista, Excite y WebCrawler. Marilyn J. Rowland. <META> Tags. En: *Beyond Book Indexing*. Diane Brenner and Marilyn Rowland, eds. Medford: Information Today, The American Society of Indexers, 2000, p. 72.
- Medeiros señalaba que sólo Altavista, Go, Hotbot e Inktomi, tienen en cuenta las metaetiquetas. Norm Medeiros. XML and the Resource Description Framework: The Great Web Hope [documento HTML]. *Online*, September 2000. Disponible en: <http://www.onlineinc.com/onlinemag/OL2000/medeiros9.html> (consultado el 26 de agosto de 2000).
- Mettrop, en el estudio que hemos citado, en el que evaluaba las fluctuaciones de acceso a los documentos a través del análisis de 13 motores de búsqueda, dice que las etiquetas de título las soportan los 13 motores analizados (Cfr. *supr.* Nota 12); la metaetiqueta `keywords`, la utilizan Altavista, HotBot, Ilse, InfoSeek, MSN, Snap y Vindex; la etiqueta `description`: Altavista, HotBot, Ilse, InfoSeek, MSN, Search.nl y Snap; la metaetiqueta relativa al autor: HotBot, MSN Snap; e incluso señala que indizan también los comentarios <!. . .> el buscador Vindex (es interesante esta apreciación de Mettrop, en el sentido de que, en muchas ocasiones, estas etiquetas incluyen metainformación interesante en páginas Web sencillas que no está reflejada de ninguna otra forma en el recurso; también hace un análisis de los sistemas que indizan los textos alternativos (atributo `alt` de HTML) que puede ser útil para la indización de imágenes, siempre en el caso también de que éstas tengan una descripción textual como atributo (Cfr. WAI, Capítulo 6, nota 45), los sistemas que indizan `alt` son Altavista, EuroFerret, Lycos y Search.nl. Wouter Mettrop. *Op. cit.*, p. 645, tabla 8.

(Altavista, HotBot, MetaCrawler, Inktomi, Infoseek, entre otros), lo hacen sobre valores básicos (*Description* y *Keywords*) de las metaetiquetas HTML, o a lo sumo la metaetiqueta *author* u otras partes del documento HTML que puedan entenderse en cierto sentido, como metainformación (el texto de los comentarios o las descripciones textuales alternativas en el elemento *IMG*⁶⁷).

Estas metaetiquetas HTML se usan aproximadamente en un 34% de los sitios Web⁶⁸, pero es notorio que los *search engines* de este tipo (Altavista, HotBot, etc.), aunque en principio las indiquen y estén preparados para basar la indización en ellas, tienden a obviarlas para sus cálculos de relevancia. Esto se debe a que, salvo en los casos en que las páginas están en servidores "de confianza", la metainformación la utilizan los proveedores de contenidos para el *spamming*⁶⁹, de tal forma que puedan conseguir mejores posiciones en los ranking de relevancia de los buscadores genéricos y aumentar su visibilidad en la Web⁷⁰. Así lo dicen explícitamente algunos de los grandes buscadores (Altavista, Northern Light, por ejemplo) al definir sus políticas de indización y relevancia:

- Altavista⁷¹ explica que las *metatags* no tienen preferencia en la indización al considerarlas como una oportunidad para el abuso y el *spamming*. Lo que

⁶⁷ Wouter Mettrop. *Ibid.*

⁶⁸ Steve Lawrence, C. Lee Giles. Accessibility of... *Op. cit.*, <http://www.wwwmetrics.com>

⁶⁹ *Vid.* Glosario.

⁷⁰ Esta circunstancia la tienen en cuenta tanto los editores de metaetiquetas como las iniciativas, especificaciones, etc. de metadatos de propósito general como el DC (recordemos herramientas como Hi-Verify, *Vid.* 4.3.2.2., Fig. 39). DCMI. *FAQ*: <http://dublincore.org/resources/faq/#whatsearchenginingsupport>, o autores particulares que, como Weinheimer, han tratado profundamente el tema. James L. Weinheimer. *Op. cit.*, p. 20. *Cfr.* Capítulo 4, nota 95.

⁷¹ How Altavista Works: Metatags: http://help.altavista.com/adv_search/ast_haw_metatags y Combating Spam: http://help.altavista.com/adv_search/ast_haw_spam En 1998 Altavista indizaba más de 1024 caracteres en las etiquetas correspondientes a *keywords* y *description* (*Cfr. sup.* Nota 65. Traugott Koch. Documentation to: Metadata... *Op. cit.*, <http://www.ub2.lu.se/~traugott/Mdsearch-docu.htm>), sin embargo en la actualidad no tienen ninguna preferencia y se

[cont.]

más le importa a los usuarios de Altavista —continúan— es el contenido visible en las páginas Web, no las notas orientadas al mercado y añadidas en las metaetiquetas. Por ello, para el cálculo de la relevancia tienen en cuenta el título de la página y las primeras líneas que, según el criterio de este motor de búsqueda, estarán probablemente relacionados con el contenido principal de la página.

- Northern Light⁷² declara que utiliza varios factores para calcular la relevancia de los resultados como el número de ocurrencias del término de búsqueda en la página o que dicho término aparezca en el título. Sin embargo afirman que *no asigna ninguna relevancia particular a las palabras que contienen las metaetiquetas, ni tampoco las utiliza para las descripciones de la lista de resultados*, es decir, para el resumen que aparece asociado a cada ocurrencia.
- Excite⁷³ sin embargo, aunque no las utiliza para el cálculo de la relevancia, sí toma como preeminente la etiqueta `<META NAME="description">` para ofrecer el resumen de los resultados.
- También es habitual entre estos sistemas que obvien las metaetiquetas para el cálculo de la relevancia (aunque sí las tengan en cuenta como parte más del texto), porque ordenan los resultados por la relación de una página/sitio Web con otras, esto es lo que se denomina tecnología *page-rank* y es el criterio que utilizan algunos *search engines*, como Google.

indizan, a lo sumo, como cualquier otro componente textual del contenido de un recurso Web. Esto demuestra la evolución de la indización de los sistemas de recuperación globales de la Red en cuanto a los metadatos HTML más básicos, que hacen pensar también en otras tendencias o fórmulas de recuperación que protegidas del riesgo de *spamming*.

⁷² Northernlight (General Faq). How does Northern Light Determine the Relevancy of a Document or Web Site to a Given Search?: http://www.northernlight.com/docs/gen_help_faq.html#q14

⁷³ Understanding Metatags: http://www.excite.com/info/getting_listed/meta_tags

En cuanto a los *search engines* que indizan y recuperan según otros modelos de propósito general, como el DC embebido en metaetiquetas del código fuente HTML, sólo un 0,3% de las páginas de la Web utilizaban en 1999 metadatos según la semántica del DC⁷⁴. Tony Gill⁷⁵, opinaba categóricamente en 1997, que la principal causa de la no-implantación de este formato, se debía que la mayor parte de los vendedores de motores de búsqueda no lo soportaban y añadía además, que era improbable que empezaran a soportarlo hasta que hubiese una base de uso bastante grande (la masa crítica a la que nos referíamos en el capítulo anterior) como para que valga la pena incluirlos. Sin embargo, esta situación ha cambiado en los últimos años. Según una encuesta realizada por la DCMI en la lista de distribución dc-general⁷⁶, los *search engines* que pueden basar su recuperación en el Dublin Core son: Ultraseek (Inktomi), Swish-E⁷⁷, Microsoft's Index Server, Blue Angel Technologies MetaStar, y Verity Search97 Information Server. No obstante, estos motores no suelen ser la base de sistemas de recuperación global, sino más bien son *engines* aplicables y aplicados generalmente a colecciones o bibliotecas digitales. Por ejemplo, MetaStar es una de las aplicaciones que hemos denominado Sistemas Integrados de Gestión de Documentos entendidos como Objetos (SIG-DLOs) y por ende, su motor de búsqueda está dirigido a la recuperación de información en bibliotecas digitales. La utilización de este tipo *search engines* no tiene ningún

⁷⁴ Steve Lawrence, C. Lee Giles. Accessibility of... *Op. cit.*, <http://www.wwwmetrics.com>

⁷⁵ Tony Gill. The Dublin Core Metadata Element Set... *Op. cit.*, p. 46. De todas formas, suscribimos la opinión de Gill, y pensamos que los metadatos serán un método efectivo para la descripción y recuperación de información en el ámbito general de la WWW, no sólo si se convierten en estándares nacionales o internacionales, sino también si se aplican consistentemente por los creadores de recursos Web y las aplicaciones se fundan en ellos. Cfr. 6.1, 6.3.

⁷⁶ DCMI FAQ, *Op. cit.*, <http://dublincore.org/resources/faq/#whatsearchenginessupport>

⁷⁷ Este *search engine* se utiliza para recuperar información en colecciones digitales o servicios de búsqueda basados en la tendencia semántica, como por ejemplo LII (*Vid. infr.*) Un listado de los sitios que lo utilizan se puede consultar en: <http://swish-e.org/sites.html>

sentido si no hay control, como ocurre en la generalidad de la WWW, sobre la asignación de metadatos en la edición de DLOs.

Aunque hemos reflejado someramente la realidad de los motores de búsqueda actuales haciendo hincapié en su relación con los metadatos, bien aplicados a toda la WWW como servicios públicos genéricos de recuperación de información (Altavista, Northern Light, etc.), bien como software comercial desarrollado para aplicar a distintos sitios o servicios en red (Verity, MetaStar, etc.), existen además múltiples investigaciones en torno a la aplicabilidad de metadatos para optimizar la RI global distribuida en la WWW.

Así por ejemplo, Jenkins y otros investigadores⁷⁸ crean un sistema que utiliza metadatos para la RI. Su propuesta se sustenta en un clasificador automático basado en la DDC, como un componente más de un motor de búsqueda distribuido, que trata de aunar las ventajas de la indización automática en texto completo, de la mayoría de los buscadores, con la precisión de la intervención manual en algunos de los directorios clasificados. Se trata de un sistema orientado a objetos, hecho en Java, que recupera documentos HTML, analiza el contenido y le asigna la correspondiente clasificación calculando el *coeficiente de Dice*⁷⁹; posteriormente representa los elementos de metadatos en el modelo de datos RDF, definidos utilizando un esquema RDF y automáticamente generados en sintaxis XML/RDF.

Por otra parte, Lowe describe un sistema prototipo alternativo para definir el contexto de las búsquedas y aumentar la relevancia de los resultados sin reducir la precisión. Argumenta para ello que los metadatos como RDF requieren mucho rigor para mantener su uso efectivo en la Web, sin embargo *su sistema basado en enlaces*

⁷⁸ Charlotte Jenkins, et al. *Op. cit.*, <http://www8.org/w8-papers/2c-search-discover/automatic/automatic.html> Cfr. 4. 2.

⁷⁹ El coeficiente de Dice es un elemento considerado tradicionalmente en el ámbito de la recuperación de Información. Así, Rijsbergen, por ejemplo, lo consigna en su libro sobre RI de los 70's. C. J. van Rijsbergen. *Op. cit.*, <http://www.dcs.glasgow.ac.uk/Keith/Chapter.3/Ch.3.html>

hipermediales es un mecanismo estructural inherente, que se genera como parte natural del desarrollo de contenido para la Web, y que puede utilizarse para determinar automáticamente el contexto⁸⁰. Según esto, este autor entiende los documentos enlazados como una forma implícita de metadatos.

Por otra parte, herramientas (*search engines*) que han tenido mucho éxito en los últimos años para la indización de sitios o servidores Web locales, con proyección global, como htDig o AtomZ, permiten la recuperación de información en metaetiquetas⁸¹ HTML. Asimismo se está introduciendo la gestión de metadatos como una fortaleza en el desarrollo de estos productos. Por ejemplo, AtomZ Search ha cambiado en los últimos meses⁸²: de ser una herramienta que proporcionaba soluciones de búsqueda para pequeños sitios, ha pasado a ser una herramienta empresarial con un motor que extrae contenido de metadatos. El *search engine* introduce metadatos para ayudar en el etiquetado de los resultados de búsqueda⁸³.

Los metadatos son pues, una tendencia creciente o al menos una alternativa observada para optimizar la RI de los *search engines* desarrollados para la información en red, pero existen también otras tendencias de sistemas que abogan

⁸⁰ David Lowe. Improving Web Search Relevance: Using Navigational Structures to Provide a Search Context [documento HTML]. En: *AusWeb2K: Australian World Wide Web Conference (6. 2000. Cairns)*. Disponible en: <http://ausWeb.scu.edu.au/aw2k/papers/lowe/paper.html> (consultado el 23 de octubre de 2000).

⁸¹ Por ejemplo htDig considera para la indización, las etiquetas META: `htdig-keywords`, `htdig-noindex`, `htdig-email`, `htdig-notification-date`, `htdig-email-subject` robots, `keywords` y `description`. Y permite configurar el peso asignado a las distintas etiquetas para el cálculo de la relevancia en la búsqueda, a través de las opciones de configuración `keywords_factor`, `use_meta_description` y `meta_description_factor`. Más información en: <http://htdig.sourceforge.net/meta.html> y en <http://htdig.sourceforge.net/attrs.html>

⁸² Mana Tominaga. From the Expo Floor: Atomz [documento HTML]. *Webreview*, 7 september 2001. Disponible en: http://www.webreview.com/web2001/product_news/index02.shtml (consultado el 25 de septiembre de 2001).

⁸³ Más información sobre el *Content Mining Engine* de AtomZ y sobre la interfaz de gestión de metadatos en: http://www.atomz.com/products/atomz_search/01_enterprise_search.htm#11

por otras técnicas. Así por ejemplo, algunos *search engines* admiten la expansión de la pregunta usando técnicas de procesamiento del lenguaje natural (PLN)) y el análisis semántico y pragmático. Este es el caso de algunas de las aplicaciones desarrolladas por MNIS (*Manning & Napier Information Services*) <<http://www.mnis.net>> que, como DR-LINK (*Document Retrieval using Linguistic Knowledge*), se basan en un sistema de procesamiento de textos inteligente que analiza el documento completo no sólo por las palabras sino también por el significado del texto. Otros sistemas basan su búsqueda en modelos de emparejamiento basados en redes neuronales como Convera <<http://www.convera.com>> (antes Excalibur RetrievalWare).

La importancia de los metadatos en estas aplicaciones de "segunda generación" que se basan más en la indización por conceptos que en la indización por palabras, la vemos por ejemplo en la tecnología desarrollada por Applied Semantics denominada CIRCA (*Conceptual Information Retrieval and Communication Architecture*). La tecnología CIRCA se basa en dos elementos fundamentales: una ontología y un motor de procesamiento lingüístico que trabaja con cualquier estructura o aplicación XML. Además de utilizar la clasificación automática, incluye un generador de metadatos que trabaja con más de cincuenta formatos de fichero diferentes. La creación de metadatos de Applied Semantics implica más que una vinculación un modelo de metainformación, la relación de los metadatos con bases de conocimiento y/o ontologías, es decir, plantea una visión de metadatos orientada a las materias.

Este generador de metadatos implica la automatización del proceso de hacer la información visible a través de la gestión del contenido, permitiendo además incorporar vocabularios controlados *ad hoc* que se integran con la estructura de la metainformación. A pesar de que CIRCA es, desde enero de 2001, una tecnología propietaria de Applied Semantics dirigida sobre todo a portales institucionales o a la gestión del conocimiento en Intranets, tiene su origen en el Instituto de Tecnologías de la Información de la Universidad de California en Los Ángeles, donde Adam Weissman y Gil Elbaz desarrollaron a finales de 1999 el SRII que conocemos como Oingo <<http://www.oingo.com>>. Oingo indiza todo el directorio de ODP y permite

tanto la navegación como la búsqueda por palabras o frases (aparentemente también por lenguaje natural). Los resultados permiten refinar la búsqueda por significados más específicos de la palabra (Fig. 55⁸⁴). Una de las mayores ventajas de trabajar desde un léxico de significados es que los usuarios pueden hacer una búsqueda conceptual⁸⁵.

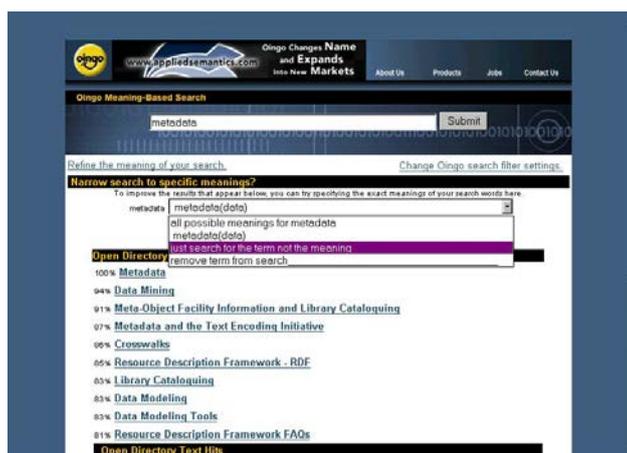


Fig. 55. Oingo (resultados de búsqueda)

Finalmente debemos mencionar los metabuscadores, que también pertenecen a los sistemas basados en una perspectiva práctico-tecnológica de la RI. Este tipo de SRII también reciben el nombre de buscadores en paralelo (e incluso, aunque con no demasiada precisión, multibuscadores). Son sistemas que permiten una localización de recursos múltiple y simultánea en varios motores a la vez, dando lugar a búsquedas aún más exhaustivas al cruzar la cobertura de varios robots de la Web y

⁸⁴ La Fig. 55 muestra una búsqueda en Oingo por la palabra "metadata". Los resultados son una serie de categorías de Open Directory (ODP) y además la posibilidad de refinar la búsqueda a "todos los significados de la palabra Metadata", "Metadata (data)", "buscar sólo por término, no por significado" (los buscadores como Altavista, Google, etc. buscan siempre por término, Oingo busca, por defecto, por significado) o "eliminar un término de la búsqueda".

⁸⁵ Paula J. Hane Beyond Keyword Searching: Oingo and Simpli.com Introduce Meaning-Based Searching [documento HTML]. *Information Today*, 20 December 1999. Disponible en: <http://www.infotoday.com/newsbreaks/nb1220-2.htm> (consultado el 23 de octubre de 2000). Esta opinión de Hane sobre Oingo se resume en el lema del buscador: *We know what you mean* (sabemos lo que quieres decir).

eliminando los URLs duplicados⁸⁶. La relación de estos sistemas con los metadatos, a pesar de su nombre "*metabuscadores*", es la que tengan cada uno de los motores con los que trabaja, es decir, la política de indización que tenga cada uno de ellos con relación a las metaetiquetas. Uno de los aspectos más significativos que debemos señalar en virtud de estos sistemas y los metadatos, es el conflicto terminológico que evocan. Los investigadores australianos del *Digital Learning System*⁸⁷ defienden que el término metabuscador (*metasearching site*) se refiere (o debería referirse) a los motores de búsqueda que utilizan metadatos. Para estos autores, el resto de sistemas que realizan la búsqueda en múltiples motores de búsqueda, incluidos los basados en el protocolo Z39.50, e incluso sistemas experimentales que pertenecen al ámbito de

⁸⁶ A este tipo de sistemas de búsqueda en paralelo pertenecen, entre otros:

- DigiSearch: <http://www.digiway.com/digisearch>
- Ixquick: <http://www.ixquick.com>
- Kartoo (francés): <http://www.kartoo.net>
- Mamma: <http://www.mamma.com/psearch.html>
- MetaCrawler: <http://www.metacrawler.com>
- Profusión: <http://www.profusion.com>

⁸⁷ Albert Ip, et al. *Metasearching or Megasearching: Toward a Data Model for Distributed Resource Discovery*. En: *e-Education: Challenges and Opportunities. Hong Kong Web Symposium (5. 1999. Hong Kong)*. Fred Castro, Robert Lai and Sr. Margaret Wong, eds. Hong Kong: Social Sciences Research Centre, 1999, p. 65-82. También disponible en Internet en: <http://www.dls.au.com/metadata/DataModel.html> (consultado el 31 de octubre de 2000). A este conflicto terminológico se nos ocurre darle una explicación de orden histórico y lógico. La diferenciación que hacen estos autores parece razonable; llamar metabuscador a los sistemas que realizan las búsquedas basándose en metainformación es si no acertado, al menos útil desde el punto de vista mnemotécnico. Sin embargo, tal y como funciona la creación lingüística para designar nuevas realidades, cuando Erik Selberg creó Metacrawler en abril de 1995, tan sólo hacía un mes del primer *workshop* del Dublin Core y aunque existían modelos de metainformación anteriores, *V. gr.* SOIF, IAFA e incluso TEIH, éstos aún no habían alcanzado la importancia y la trascendencia actual en la recuperación de información. Por ello Selberg denominó a su sistema de búsqueda Metacrawler sin más dilación; si a los motores de búsqueda se les denominaba *crawlers* (orugas), era lógico que aplicase el prefijo meta- (*Cfr.* 3.1.) para designar al sistema de búsqueda que basaba su funcionamiento buscando *a través* de varios motores, y por extensión se denominarían metabuscadores a todos aquellos que utilizasen esta tecnología. Por ello creemos que es difícil que la diferenciación que proponen estos autores australianos prospere y se denominen de forma generalizada metabuscadores, aquellos sistemas que se basen en metadatos.

la investigación, como HotOil⁸⁸, deben denominarse megabuscadorees (*megasearch sites*) o pasarelas de búsqueda (*search gateways*⁸⁹).

Todos los SRII que hemos señalado hasta ahora (directorios, buscadores, metabuscadores) e incluso otras herramientas cliente como los agentes o *SearchBots*⁹⁰, han servido y sirven, a distintos niveles, para la recuperación de información en Internet. Sin embargo, estos sistemas de indización y búsqueda tienen muchas limitaciones como servicios de recuperación. A pesar de su infraestructura técnica (bien basada en la potencia, bien en la semántica) y de su interfaz de usuario amigable, adolecen todavía, en la mayoría de los casos, de una estructuración de los datos, es decir de metadatos sobre los objetos de información, que fundamenten los *search engines* aplicables.

Tendencia bibliotecaria de carácter semántico

Ante el complejo y vasto entorno de la RI en Internet, además de las tendencias que, de forma general, hemos apuntado hasta ahora, también tenemos que considerar, iniciativas de recuperación de información globales, si bien no podemos considerarlas bibliotecas digitales en sentido estricto, sí responden a criterios bibliotecarios de organización y recuperación. La primera tendencia que podemos recordar en este sentido es la concepción de la búsqueda en Internet basada en la formación de catálogos o bases de datos bibliográficas convencionales. A este grupo pertenecerían proyectos como InterCat de OCLC, del que ya hemos hablado, que se

⁸⁸ HotOil es un motor de búsqueda prototipo desarrollado por el del DSTC (*Distributed Systems Technology Centre*) que distribuye las sentencias de búsqueda y recoge los resultados de bases de datos distribuidas. Para una información exhaustiva sobre este proyecto Vid. Nigel Ward. *Experiences with Distributed Searching* [documento PDF]. En: *VALA Biennial Conference and Exhibition (10. 2000. Melbourne)*. Disponible en: <http://www.vala.org.au/vala2000/2000pdf/Ward.PDF> (consultado el 26 de agosto de 2000).

⁸⁹ Cfr. 8.2.2., Glosario, *subject gateway/information gateway*.

⁹⁰ Vid. Glosario.

basa en la asunción de que los recursos electrónicos pueden catalogarse utilizando las reglas de catalogación y el formato MARC, con la aspiración de crear una gran base de datos en línea que permita una recuperación de calidad, facilitando incluso la consulta múltiple de bases de datos distribuidas a través de los protocolos Z39.50 o STARTS. Además de la creación de un catálogo, el objetivo global de InterCat era también *evaluar la eficacia de la utilización del USMARC para facilitar la descripción, el lugar y las condiciones de acceso para los documentos accesibles a distancia*⁹¹. Si entendemos el formato MARC como un modelo de metainformación, estos sistemas estarían basados en metadatos de la banda 3 de Dempsey y Heery. Pero tanto InterCat, como otros proyectos semejantes, demostraron la imposibilidad de enfrentar el problema de la recuperación en la Web con criterios MARC⁹².

Otra iniciativa de cariz documental para enfrentar el problema de la RI en la WWW, es la aplicación de clasificaciones bibliotecarias para crear índices o directorios de recursos, que permiten recuperar información *full-text* a través de la navegación o *browsing* de valor añadido. Podemos mencionar algunos servicios paradigmáticos de esta tendencia, como CyberStacks <<http://www.public.iastate.edu/~CYBERSTACKS>>. Surgido en 1998, se autodefine como una colección centralizada, integrada y unificada de recursos Web significativos, organizados usando el esquema de clasificación de la *Library of Congress*. Este servicio permite navegar por la información ordenada por materias, e identificar los recursos supuestamente relevantes asociados a la descripción temática que mejor caracterice su contenido y cobertura. La diferencia de este directorio con otros más genéricos, como Yahoo, es que la selección de recursos incluidos en el índice se hace según criterios estrictamente bibliotecarios extraídos de los principios establecidos en 1992 por el Comité para la Evaluación y Desarrollo de Colecciones de Referencia de la ALA, que son, entre otros: la adecuación de la información, la claridad de la presentación, la

⁹¹ Vianne Sha. Cataloging Internet Resources: the Library Approach. *Electronic Library*, 1995, vol. 5, nº 13, p. 467-476. Citado en: Roman S. Panchyshyn, France Bouthillier. *Op. cit.*, p. 140.

⁹² *Vid.* 4.1.

unidad en el almacenamiento, el nivel de actualización y las referencias favorables. Por otra parte, la diferencia fundamental con otros directorios cuyo origen también es de índole bibliotecario, como Digital Librarian⁹³, es que la clasificación no es *ad hoc* sino una clasificación preestablecida como la LCC.

Existen además servicios que combinan clasificaciones tradicionales con la búsqueda por campos o por metainformación asignada *a posteriori*, que les sitúa muy próximos a la consideración de biblioteca general digital distribuida o *multi-subject gateways*⁹⁴ e incluso para algunos, *information gateways* en sentido amplio⁹⁵. Los sistemas más representativos a este nivel son:

- BUBL <<http://bubl.ac.uk>>, un catálogo de más de 12.000 recursos de Internet seleccionados y organizados según la DDC. El resultado un esquema de clasificación jerárquico donde los ítems de contenido similar aparecen unos cerca de otros por proximidad semántica. Combina la navegación por las categorías de Dewey o un sistema de clasificación propio denominado 5:15 y basado en la LCSH, con la búsqueda que puede acotarse al nombre del recurso, resumen, autor, tipo de autor, materia, clasificación DDC, tipo de recursos, ubicación y fecha, esto es, permite una búsqueda por campos, muy parecidos a los incluidos en los elementos de metadatos del Dublin Core (DCMES).
- INFOMINE <<http://infomine.ucr.edu>> Iniciado en 1994 por las bibliotecas de la Universidad de California, es un sistema de búsqueda y recuperación da acceso más de 20.000 recursos de Internet, seleccionados y organizados por materias según los encabezamiento de la Biblioteca del Congreso (LCSH). También permite buscar por título, materia, palabra clave, autor, descripción o sobre el

⁹³ Cfr. *supr.*

⁹⁴ Así se clasifica por ejemplo en el servicio PINAKES de la Biblioteca de la Universidad de Heriot-Wat: <http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html>

⁹⁵ Cfr. Glosario, *subject gateway*.

texto completo de los documentos registrados en su base de datos. Los recursos seleccionados se indizan en una base de datos SQL y el *search engine* asociado a ella, localiza los términos de búsqueda entre las descripciones generadas por los bibliotecarios.

- Librarians Index to Internet <<http://lii.org>>. Es el proyecto más antiguo que tiene su origen en un fichero de *bookmarks* gopher de la bibliotecaria Carole Leita. En la actualidad y desde 1997 es un directorio comentado de 8.600 recursos de Internet, seleccionados y evaluados por bibliotecarios y dirigido a los usuarios de bibliotecas públicas. Permite tanto la navegación como la búsqueda acotada a campos (materia, título, descripción, autor, editor, URL, indizador, categoría o todos los campos), operadores booleanos, etc. Para las búsquedas en su base de datos utiliza el *search engine* Swish-E, que como ya hemos dicho, permite la búsqueda en metadatos DC⁹⁶.
- Internet Scout Project (ISP) <<http://scout.cs.wisc.edu/scout>> es un sistema creado por la NSF (*National Science Foundation*) en el Departamento de Informática de la Universidad de Wisconsin-Madison, para la recuperación de información de calidad en Internet en el contexto de la educación superior americana. Scout Report Archives <<http://scout.cs.wisc.edu/archives>> es un servicio de búsqueda basado en proyecto de prueba (1996-2000) denominado Signpost que se realizó para demostrar que los recursos de Internet se podían catalogar, clasificar y organizar utilizando estándares consolidados en el mundo bibliotecario, tales como los encabezamientos de materia o la clasificación de la Biblioteca del Congreso (LCSH, LCC), en coordinación con estándares de metadatos como el Dublin Core. Además han creado un *scheme* de clasificación específico denominado CYRUS (*Classify Your Resources Using Scout*) que, a fecha de finalizar la redacción de este trabajo, aún no se había hecho público, sin embargo sabemos que incluirá veintidós clases principales en las tres materias principales

⁹⁶ Cfr. *supr.* Nota 77.

de ISP. Uno de los primeros frutos de Singpost fue la creación de la red Isaac que trataba de aunar la búsqueda en pequeños y diversos directorios surgidos en Internet, a través de una única interfaz, basándose para ello en el modelo de metadatos del Dublin Core y en protocolos sencillos como CIP y LDAP.

Hoy en día el ISP se ha incorporado IMesh (*International Collaboration on Internet Subject Gateways*) un proyecto internacional financiado por la DLI2 donde participan, además de ISP, el ILRT y UKOLN, en el desarrollo de un conjunto de herramientas para la gestión de metadatos que integren distintas visiones de las bibliotecas virtuales temáticas, posibilitando acceder a los datos en entornos Z39.50. Asimismo se proponen el estudio de los problemas de la provisión de servicios dirigidos a un usuario final en redes internacionales. La evolución del ISP le convierte una pasarela de materia como las que estudiaremos en el capítulo siguiente dentro del modelo de biblioteca digital distribuida⁹⁷.

- WWW Virtual Library <<http://vlib.org>>. Aunque formalmente a este proyecto se le denomina biblioteca virtual, se trata de un directorio de búsqueda general fundado en 1991 por Tim Berners Lee. La filosofía de creación de este proyecto era semejante a la que hoy protagoniza hoy al ODP, la creación de un directorio constituido por humanos donde cada uno selecciona y categoriza distintos recursos en la materia en la que es especialista. Igualmente existe una base de datos central que sólo permite la búsqueda cualificada sobre el título. A diferencia de ODP, esta base de datos no está basada en RDF, ni en ningún otro modelo de metainformación. Además se puede buscar en cada una de las categorías según los métodos y *search engines* (htDig, etc.) que haya aplicado la persona encargada de seleccionar los recursos de ella.

Todos estos sistemas de búsqueda están basados en la tendencia teórico-semántica de la RI y en criterios bibliotecarios no-automáticos de selección,

⁹⁷ Cfr. 8.2.2., 8.3.2. IMesh: <http://www.dli2.nsf.gov/internationalprojects/JISC/IMESHtoolkit.pdf>

indización y organización de la información, sin embargo, es preciso que maticemos su posición en virtud de los metadatos de una forma clara distinguiéndolos, por un lado de los motores de búsqueda globales, y por otro de las bibliotecas digitales:

- a) Comparten con los motores de búsqueda el objetivo de indizar y representar el contenido global de la Web en una base de datos, y que esa representación es posterior a la edición de los DLOs. Comparten con las *subject gateways* o bibliotecas virtuales temáticas, el concepto de catálogo, y de "colección digital" aunque no se basan estrictamente en metadatos estructurados.
- b) La diferencia fundamental de estos sistemas con los grandes motores de búsqueda es que no se centran en la recuperación *full-text*, su *search engine* opera frente a un catálogo o directorio. No existe robot que recorra la Red de forma automática recopilando enlaces o recursos de interés, sino que la base de datos se realiza de forma manual, tanto desde el punto de vista de la selección como de la descripción. Al igual que los *search engines*, y a pesar de su sesgo académico (Infomine y Bubl), tienen un carácter enciclopédico (pretenden cubrir la información global de la Red). Combinan el modelo de organización y búsqueda de información de Yahoo⁹⁸, con las técnicas bibliotecarias tradicionales.
- c) Es su afán de ser sistemas de recuperación de información "cualificada" global (de toda la Web), lo que los diferencia de las bibliotecas digitales. Su crecimiento y cobertura, al igual que en los *search engines* automáticos, aumentará al mismo tiempo que crezca la WWW.

⁹⁸ Este modelo de organización de la información recibe denominaciones como *portal style navigation* o *portal-style interface (such as Yahoo)*. Cfr. Capítulo 4, nota 171.

- d) A pesar de que estos SRII se denominan con frecuencia bibliotecas virtuales⁹⁹, digitales o electrónicas, la diferencia fundamental entre ellos y las *subject gateways* —que consideraremos en el capítulo siguiente, como un tipo especial de biblioteca digital¹⁰⁰—, es que no son sistemas especializados, sino que pretenden dar acceso a todas las disciplinas.
- e) El sesgo eminentemente bibliotecario que confluye en estos SRII, nos lleva a comparar también estas tentativas "catalográficas"¹⁰¹ con proyectos de catalogación propiamente dicha de recursos electrónicos (v. gr. InterCat). La diferencia fundamental entre estos sistemas e InterCat, es que se basan en el formato MARC. Constituyen pues, una respuesta bibliotecaria situada entre el control bibliográfico exhaustivo y tradicional (catalogación-MARC) y la proyección de la nueva Web Semántica.

Sistematizaremos ahora todas las reflexiones que hemos pronunciado sobre las tendencias de los de SRII en la tabla 14. Incluiremos sólo aquellos SRII que hemos mencionado de manera ilustrativa en la exposición, conscientes de que existen miles de sistemas de búsqueda y recuperación de información. Nos basamos en el criterio de los metadatos para la búsqueda por palabras clave, partiendo de la tipificación de los metadatos en bandas de Depsey y Heery¹⁰², y en los sistemas de clasificación de

⁹⁹ Por ejemplo, Infomine se define como "*virtual library*". Steve Mitchell, Margaret Mooney. INFOMINE: A Model Web-based Academic Virtual Library [documento HTML]. *Information Technology and Libraries*, March 1996, vol. 15, nº 1. Disponible en: <http://infomine.ucr.edu/pubs/italmine.html> (consultado el 8 de agosto de 2000). Y también diversos autores reconocen este sistema dentro de las bibliotecas digitales, por ejemplo: David Stern. New Search and Navigation Techniques in the Digital Library. En: *Digital Libraries: Philosophies, Technical Design Considerations and Example Scenarios*. David Stern, ed. New York, London, Oxford: The Haworth Press, 1999, p. 77.

¹⁰⁰ Vid. 8.2.2.

¹⁰¹ En tanto que utilizan la catalogación podríamos decir que son SRII basados en metadatos, ya que según explicamos en el capítulo 3 (3.2), los registros catalográficos están constituidos por metainformación. Sin embargo, en el sentido estricto que damos en esta tesis doctoral a los metadatos, no pueden considerarse como tales.

¹⁰² Cfr. 3.5, Tabla 8.

materias que se utilizan para la navegación o *browsing* de valor añadido. A pesar de la dificultad de expresar categóricamente los fundamentos de estos sistemas, la tabla 14 recoge, por orden de aparición en el texto (excluyendo los proyectos de investigación citados) una muestra del complejo espacio de la RI en la Web, clasificando los servicios y/o sistemas que hemos citado en este trabajo para esclarecer las tendencias que señalamos.

La interpretación de los signos que aparecen es la siguiente:

- : implica extracción o asignación automática
- : quiere decir que el trabajo es manual, realizado por humanos.
- : esta combinación de símbolos simboliza la asignación manual de metadatos (en todos los sistemas reflejados, realizada en modo *post-publicación*), pero que al mismo tiempo tienen un *search engine* que indiza basándose en ellos. Son los que se denominan de forma general, *Catalogue-based search engines*, es decir motores de búsqueda basados en un catálogo y no en todas las páginas recabadas por un robot.
- El texto en **sombra** quiere decir que son *search engines* propiamente dichos, que a su vez (como en el caso de Swish-E) constituyen algunos de los servicios que se señalan (p. ej., Swishe-E es el *search engine* empelado en el servicio LII).
- En el caso de las metaetiquetas (*pseudo-metadatos* banda1), sólo se han señalado cuando el sistema las tiene en cuenta para el cálculo de la relevancia en las búsquedas, no si las indiza como cualquier otro elemento de texto.

Tabla 14. Características y tendencias de los SRII mencionados

	METADATOS				OTRAS	CLASIFICACIÓN			
	No <i>Full-text</i>	Banda 1 MetaTag	Banda 2	Banda 3		No	<i>Ad hoc</i>	Origen bibliot.	Bibliot. (DDC, etc.)
Yahoo						/ 			
Digital Librarian									
ODP				 RDF		/ 			
Altavista									
HotBot									
MetaCrawler									
Inktomi (Ultraseek)									
Northern Light									
Excite									
Google						/ ODP			
Swish-E									
MS Index Server									
MetaStar (Blue Angel)									
Verity Search									
HtDig									
AtomZ									
Dr-Link									
Convera									
Oingo / AppliedSem.							 tesauros		
HotOil									
InterCat				 MARC					
CyberStacks									

	METADATOS			OTRAS	CLASIFICACIÓN				
	No <i>Full-text</i>	Banda 1 MetaTag	Banda 2	Banda 3	PLN, etc.	No	<i>Ad hoc</i>	Origen bibliot.	Bibliot. (DDC, etc.)
Multi Subject gateways – Information gateways de cobertura global									
BUBL									
INFOMINE									
LII	 Swish- E	 Swish- E	 Swish- E						
WWW VL									
ISP								 CYRUS	

Tabla 14. Características y tendencias de los SRII mencionados

Los metadatos, bien creados dentro del propio documento por los autores del DLO de tal forma que permitan crear *search engines* que extraigan la información de ellos, bien creados *a posteriori* formando "catálogos de propósito general", como ODP o los catálogos de índole más bibliotecaria como BUBL, son una tendencia creciente en la concepción de sistemas de búsqueda global para la WWW.

En 1997, Younger¹⁰³ reconocía que los aspectos más importantes para construir un sistema coherente de acceso bibliográfico para los usuarios de la Web eran: *los metadatos, los objetivos de la catalogación bibliotecaria, las estructuras de registro, nombres y direcciones persistentes para los recursos accesibles en línea y la gestión de diversos esquemas de metadatos*. No obstante es preciso apuntar otras tendencias generalizadas en la Web que afectarán sin duda al desarrollo de los metadatos como tecnología aplicable a la recuperación de información global, así como a los servicios informativos y bibliotecas digitales locales. Estas circunstancias que detectamos son:

- El fenómeno de la "portalización" de los sistemas de búsqueda, esto es, la tendencia incipiente a que se conviertan en servicios de valor añadido y en un

¹⁰³ J. A. Younger. Resources Description in a Digital Age. *Library Trends*, 1997, vol. 45, nº 3, p. 462-461. Citado en: G. G. Chowdhury. *Op. cit.*, p. 434.

punto de partida (*home* de navegador) para los usuarios. Constituyen un nuevo modelo de negocio empresarial y desde nuestro punto de vista, están retrasando el micro-pago por las búsquedas vaticinado desde hace algún tiempo por Nielsen. En los portales de información tienen también cabida los metadatos, sobre todo, formatos de sindicación de contenidos¹⁰⁴ basados en RDF como RSS, y sobre todo, la creación de metadatos dirigidos al comercio electrónico, a través de esquemas (*schemas*) como <indecs>.

- La verticalización de la información. Esta verticalización y/o especialización de los servicios de información se manifiesta tanto en aquellos dirigidos a un usuario genérico global, como en la información científico-técnica dirigida a usuarios especializados o locales. En la primera vertiente, la especialización informativa está dando lugar a *vortales* o portales verticales, y en la facción de información científica ha dado lugar a múltiples de comunidades virtuales, *subject gateways*, *clearinghouses*, etc. que aspiran a crear modelos normalizados para su información¹⁰⁵, en muchos casos, basados en metadatos.
- Por último, la tendencia más importante con relación a los metadatos y el futuro de los SRII es el desarrollo e implantación inminentes de XML para la creación de contenidos digitales, así como la tendencia a la Web Semántica basada tanto en el desarrollo de metainformación orientada al recurso, como de ontologías, etc. y de los complejos modelos de RI en desarrollo asociados a ellos.

¹⁰⁴ Vid. Glosario, *syndication/ syndication format*.

¹⁰⁵ Así por ejemplo se han creado consorcios como el OpenGIS <<http://www.opengis.org>> para la organización y modelización de los Sistemas de Información Geográfica.

7.3. Búsqueda en texto completo *versus* búsqueda en metadatos: ventajas e inconvenientes

Los mecanismos de búsqueda y recuperación de información son inseparables del mecanismo de indización, tanto para las técnicas de fichero inverso como para las que dependen de la lingüística o de otras. Por ello, plantharemos a continuación un panorama genérico a las ventajas e inconvenientes que tiene la búsqueda basada en metadatos, frente a la búsqueda en texto completo, o lo que es lo mismo, la indización (*harvesting*) basada en *datos estructurados sobre datos*¹⁰⁶, frente a la indización *full-text* o del texto completo.

En el apartado anterior de este mismo capítulo, hemos señalado las principales tendencias y características de los SRII, haciendo hincapié en que la mayoría de los *search engines* de uso general en Internet lanzan algoritmos de búsqueda contra sus bases de datos construidas de forma automática a través de una indización del texto completo de los documentos existentes en Internet, mientras que los sistemas de recuperación *de organización bibliotecaria* construyen índices o estructuras de "metainformación". Abadal y Estivil¹⁰⁷ mencionan varios problemas de los buscadores globales de la Web (que serán al mismo tiempo algunas de las razones que justifiquen la formación de bibliotecas digitales que discutiremos en el capítulo siguiente): no disponen de criterios de calidad para la selección de recursos que formarán parte de sus bases de datos; las descripciones de los recursos son muy elementales incompletas y erróneas, no presentan las opciones de recuperación de los catálogos o de las bases de datos comerciales; la unidad documental que consideran

¹⁰⁶ En esta definición implícita de metadatos, esa estructura es, según Iannella, el elemento crucial que otorga a los metadatos el límite por encima de la indización *full-text*, ya que esa estructura alude a la semántica de los metadatos. Renato Iannella. Mostly Metadata... *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA1998>. *Cfr.* Capítulo 3, nota 57.

¹⁰⁷ Ernest Abadal, Assumció Estivil. *L'accés als recursos Web des de les biblioteques* [documento PDF]. Barcelona: Consorci de Biblioteques Univeritaries de Catalunya, gener de 2000. Disponible en: <http://www.cbuc.es/5digital/ireweb.pdf> (consultado el 3 de noviembre de 2000).

es el fichero y no el recurso (el DLO); los recursos cambian de ubicación y los buscadores no los actualizan sistemáticamente; estos autores¹⁰⁸, añaden además que en muchos casos la consulta es lenta.

En una primera aproximación a los procesos seguidos por los sistemas de búsqueda que vamos a diferenciar ahora, podemos decir que:

- Los sistemas de recuperación basados en texto completo, indizan todo el documento (*full-text*) donde los términos extraídos de las páginas HTML pasan a formar parte de sus índices y la búsqueda se realiza por coincidencia exacta (*exact matching*) de los términos de la pregunta con los del fichero inverso constituido automáticamente. El principal problema de estos índices a la hora de la recuperación es que su información semántica es inadecuada o inexistente, ya que la indización suele basarse en la frecuencia de aparición de las palabras, lo que conlleva que el desfase entre la relevancia objetiva —que es la que consiguen, como vimos antes— y la relevancia para el usuario, sea muy grande.

La indización se realiza por palabras clave que, por definición, no están tomadas de un vocabulario controlado, además no tienen un contexto concreto¹⁰⁹, no existe relación con otras palabras que podrían coadyuvar (por ejemplo, por jerarquía) a localizar información. A esto hay que añadir que, salvo en algunas ciencias básicas que tienen una terminología muy estable, los términos que se utilizan distan mucho de ser precisos y de un significado unívoco.

- Los sistemas de recuperación basados en metadatos realizan una indización denominada *harvesting*. El *harvesting* se define básicamente como: *la operación inversa a la creación de metadatos, en el sentido de que una herramienta de 'harvesting' necesita disociar los metadatos en categorías de indización, y que el*

¹⁰⁸ *Ibid.*

¹⁰⁹ *Vid. infr.* Reflexiones en torno a la visión lingüística de la recuperación de información en Blair.

primer objetivo del 'harvesting' es hacer útiles los metadatos para la búsqueda¹¹⁰. Los SRII globales, basados en metadatos, se fundamentan, en la mayoría de los casos, en el modelo de indización y búsqueda distribuida¹¹¹: donde los datos de indización pueden extraerse por una aplicación informática (p. ej. Harvest¹¹²) antes de que la información se transmita a través de la Red, estos datos de indización previamente extraídos se colocan en un formato estructurado de metadatos (p. ej. SOIF, si seguimos hablando de Harvest) y se transmiten usando un protocolo de compresión de información que soporta actualizaciones cada vez mayores.

- Uno de los problemas fundamentales de los SRII que utilizan el *harvesting* como modo de indización, es la calidad de los metadatos fuente, teniendo en cuenta que la indización se puede realizar de dos formas: a) construyendo un índice general de distintas materias o dominios de la Red —es decir, un buscador global, tipo

¹¹⁰ *Metadata Workshop: Report of the Workshop Held in Luxembourg...* Op. cit., p. 10-11. Cfr. Glosario, *harvest-harvesting /gather-gathering*.

¹¹¹ Sobre el modelo de indización y recuperación distribuida, son interesantes las contribuciones de: Michael F. Schwartz, creador de Harvest, el software más extendido que se basa en estructuras de metadatos Vid. Distributed Indexing... Op. cit., <http://www.w3.org/Search/9605-Indexing-Workshop/Papers/AllPaps.doc>, p. 8. Denenberg, por su parte, habla de la búsqueda distribuida como alternativa al modelo de indización global tradicional (*full-text*), comparando ambos sistemas y señalando las limitaciones del modelo distribuido. Ray Denenberg. Op. cit., <http://lcweb.loc.gov/z3950/agency/papers/italy.html> Es importante señalar las últimas técnicas para armonizar la heterogeneidad semántica de bases de datos distintas a través de lo que se denominan *bases de datos federadas*, que responden a una concepción y arquitectura de metadatos para la recuperación distribuida. Sobre este tema, Vid. Goksel Aslan, Dennis McLeod. Semantic Heterogeneity Resolution in Federated Databases by Metadata Implantation and Stepwise Evolution. *The VLDB Journal*, 1999, nº 8, p. 120-132. Modelo semejante es el que se plantea en: Bruce Schatz, et al. Federated Search of Scientific Literature. *Computer*, February 1999, p. 51-59.

¹¹² Utilizamos aquí Harvest como ejemplo de referencia en esta descripción genérica de SRII basados en metadatos, ya que junto ROADS, han sido durante mucho tiempo las aplicaciones informáticas de dominio público más utilizadas en la indización distribuida de la WWW. Harvest se basa en una arquitectura repetida en distintos servidores Web, que consta de dos elementos principales *gatherers* (sistemas que extraen y estructuran la información) y *brokers* (sistemas que indizan y recuperan esa información extraída por los *gatherers*). Utiliza uno de los primeros formatos de metainformación muy elemental, SOIF (*Summary Object Interchange Format*). Cfr. Glosario, *harvest-harvesting /gather-gathering, broker*.

AliWeb¹¹³—; y b) construyendo índices para dominios específicos o áreas determinadas de información (por ejemplo, sólo información gubernamental basándose en un único formato de metadatos, como por ejemplo, GILS).

Por otra parte, y antes de pormenorizar las ventajas de la búsqueda basada en estructuras de metadatos que, de forma elemental, no serán otras que el intento de solución a los problemas o inconvenientes de la indización y búsqueda *full-text*, nos parece interesante partir de un argumento basado en la importancia del lenguaje en la RI. Al hablar de la importancia del lenguaje, no nos referimos sólo al idioma en que estén concebidas las páginas Web que se van a indizar, sino a que la semántica profunda de un DLO, no sólo depende de la frecuencia de aparición de un término supuestamente significativo, sino que está mediatizada por el lenguaje conceptual tácito y por el contexto temático en el que se inserta o por el ámbito de aplicación¹¹⁴. En esta línea, Blair¹¹⁵ señala que, en la recuperación basada en el texto completo, el mayor problema es que los sistemas que realizan ese tipo de indización se

¹¹³ AliWeb (*Archie-Like Indexing for the Web*), es un SRII basado en metadatos y surgido en 1993 por iniciativa de Martin Koster. En este sistema, el administrador de cada servidor crea un índice normalizado con un formato de metainformación muy básico IAFA (banda 2 según la clasificación de Dempsey y Heery. *Cfr.* Capítulo 3, nota 107) que contiene información estructurada sobre los ficheros HTML, posteriormente se registra su servicio en AliWeb. AliWeb recuperará mediante un agente automático los ficheros de índices de los servidores registrados. La particularidad de AliWeb es su alto componente de indización humana.

¹¹⁴ El contexto determina la carga semántica de un concepto y es fundamental tenerlo en cuenta para la recuperación. No tendrá el mismo significado la secuencia de caracteres "Madonna" en un documento sobre música que en uno sobre arte. No obstante, en una búsqueda a un sistema basado en la indización del texto completo, un documento relativo a la cantante se presentaría como resultado relevante (relevancia objetiva) ante una consulta que lanzara un usuario por la palabra clave Madonna para encontrar una pintura italiana de la Virgen (relevancia subjetiva).

La importancia del contexto para la creación de sistemas de búsqueda de documentos a texto completo ha sido señalada por muchos autores, por ejemplo, David Lowe. *Cfr. sup.* Nota 80. Ward y Wood, por su parte, señalan que el proceso de recuperación de información (incluso en las bibliotecas tradicionales) es una operación contextual. Destacan por ello, la dificultad que tienen los sistemas de indización y recuperación *full-text* en abstraer el contexto de la información, dificultad incrementada por la propia complejidad y heterogeneidad del espacio de información Web. Nigel Ward and Andrew Wood. *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA98>

¹¹⁵ David C. Blair. *Op. cit.*, p. 209-211.

fundamentan en la asunción de que la mera tabulación de las ocurrencias de las palabras en un texto es suficiente para determinar o predecir la materia/s de la que trata. De acuerdo con este autor, afirmamos que los documentos necesitan una estructura lógica de representación con la que puedan facilitar acceso a dichos objetos de información. Esta estructura lógica, no deriva normalmente del texto de los documentos, en parte porque la estructura lógica se basa en cómo se usan esos documentos en los contextos en los que se crearon e implementaron. De ahí la importancia de utilizar XML que señalábamos en el capítulo 5, ya que permite añadir semántica al texto, reduciendo ostensiblemente este problema.

Estas reflexiones basadas en la visión lingüística y conceptual de la RI, justifican en cierta medida que existan distintos formatos de metadatos adaptados a cada disciplina. Además, los lenguajes (naturales) a través de los cuales se representa el conocimiento, pueden compartir el código pero no el significado, y en muchos casos, no comparten ni tan siquiera el código, porque pueden tener formas de expresión multimedia distintas, que conviven en una misma colección, e incluso en un mismo DLO. En estos casos la indización conjunta, basada en frecuencia de aparición textual es, desde cualquier punto de vista, deficiente.

Además debemos matizar que la delimitación entre sistemas basados en texto completo y los sistemas basados en metadatos no es tan diáfana como se esboza aquí. En consonancia con lo expresado en el apartado anterior, muchos de los SRII utilizan ambas técnicas e incluso técnicas intermedias de utilización del texto completo para la extracción automática de metadatos. En esos sistemas mixtos sería más fácil pensar, parafraseando a Aubrie¹¹⁶, en una entidad metadatos-datos que se concibe perfectamente en el marco de una mejor gestión de los vínculos existentes entre un documento y la referencia al mismo. Esa entidad mixta metadatos-datos permite asimismo una indización muy sencilla del documento: indización sobre la propia referencia incluida en el documento (metadatos) al mismo tiempo que sobre el texto

¹¹⁶ Claude Aubrie. *Op. cit.*, p. 291.

completo o mejor aún, sobre un documento íntegro estructurado. Sin embargo, y para dejar patentes las ventajas de la búsqueda basada en metadatos, seguiremos tomando individualmente ambos sistemas de búsqueda para reflejar una mayor claridad expositiva a este respecto.

Los motores de búsqueda que utilizan robots para recopilar el contenido de las páginas Web, a pesar de que son herramientas muy útiles para la ubicación de recursos en la Red, presentan una problemática que Roszkowski y Lukas¹¹⁷ dividen en dos categorías:

- Por un lado, la dificultad en la formulación de preguntas que sean lo suficientemente discriminatorias para que devuelvan un número “razonable” de ocurrencias. Normalmente cuando se lanza una consulta que contiene palabras comunes, suelen emitir miles de respuestas relevantes (objetivamente), es decir, lo que tradicionalmente en RI se clasificaría como sistemas muy exhaustivos pero poco precisos. Por otro lado, el usuario cuenta con poco más que una lista de cientos o miles de URLs cuya relevancia subjetiva sólo se podrá averiguar accediendo a cada una de ellas y discriminando lo que realmente le resulta útil a sus expectativas¹¹⁸.
- Por otra parte, el segundo problema de SRII basados en el texto completo, es el control de calidad de los resultados. La mayor parte de los motores de búsqueda no evalúan la calidad de los recursos que indizan, de tal forma que aún formulando una pregunta muy específica (utilizando operadores booleanos, por ejemplo) probablemente le devolverá una lista más o menos amplia de URLs, supuestamente relevantes, desde el punto de vista formal.

¹¹⁷ Michael Roszkowski, Christopher Lukas. A Distributed Architecture for Resource Discovery Using Metadata [documento HTML]. *D-Lib Magazine*, June 1998, vol. 4, nº 6. Disponible en: <http://www.dlib.org/dlib/june98/scout/06roszkowski.html> (consultado el 18 de julio de 2000).

¹¹⁸ En algunos casos, los SRII global (V. gr. Excite), utilizan el contenido de la etiqueta <META NAME="description"> para reflejar el resumen o descripción del recurso y de tal forma anticipar así al usuario el contenido de la información Web. *Cfr. supr.* 7.2.

Sin embargo, algunos e incluso todos los resultados pueden tener un valor escaso debido a la falta de autoridad y de validez u oportunidad de la información en la materia. La contrapartida más usual a este problema es la creación de sistemas de búsqueda basados en la selección de recursos, clasificándolos incluso, con criterios bibliotecarios.

Centrándonos ahora en términos concretos de los SRII, podemos añadir que, a pesar de su rapidez de generación de índices y la rapidez de consulta, la creación automática de ficheros inversos partiendo del texto completo tiene múltiples problemas a la hora de la recuperación, debido a que¹¹⁹ :

- Los software-robot que recorren la World Wide Web para descubrir e indizar nuevos recursos, incrementan y puede saturar el tráfico en la Red. El número de estos robots, es cada vez más grande y los administradores de los sitios Web pueden impedir que indiquen sus páginas, que sigan los enlaces contenidos en ellas, o ambas. Así los resultados de las búsquedas representar sólo una parte de la información real existente.
- La indización realizada sobre el texto completo, suele fundamentarse en el recuento estadístico de las frecuencias de aparición de las palabras, entendidas éstas como secuencias de caracteres o significantes lingüísticos. No son capaces de discernir la importancia relativa de los documentos (páginas principales, páginas con información administrativa, etc.), y no identifican tampoco el contexto de la información. A pesar de que los sistemas de comprensión automatizada del lenguaje natural han avanzado mucho en los últimos años (sistemas como Dr-Link, la tecnología CIRCA de Oingo, etc., lo demuestran), aún están en desarrollo, como para basar la indización en los significados en lugar de en los significantes.

¹¹⁹ Aquí planteamos únicamente, características generales de los sistemas basados en la indización *full-text*. Para una información detallada o individualizada de cada uno de estos sistemas, *Vid.* Traugott Koch. Detailed Description of... *Op. cit.* http://www.ub2.lu.se/tk/websearch_systemat.html#detai

- En general, los sistemas basados en la indización automática *full-text*, sólo indizan texto, existiendo gran dificultad para extraer de manera automática la semántica de DLOs no-textuales (imagen digital, multimedia, etc.).

Para paliar, o al menos soslayar, los problemas de la indización automática basada en el texto íntegro existen, como venimos perfilando a lo largo de este trabajo, distintas propuestas o alternativas, algunas de ellas, de cariz bibliotecario, pero en cualquier caso, basadas en la cooperación entre los proveedores de índices y de recursos de información y los creadores de información (autores/editores). Una de estas propuestas para solucionar los inconvenientes de la búsqueda *full-text*, es tener una estructura de índice normalizada y construir un sistema bibliográfico usando metadatos asignados *a priori*, lo más cerca posible del momento de creación del documento, que luego permitirán la indización localizada y restringida a estas estructuras concretas de metainformación, en lugar de palabras clave extraídas *a posteriori* sobre el documento total. En este sentido, de acuerdo con Desai¹²⁰, argumentamos que la indización (y posterior búsqueda y recuperación) distribuida basada en metadatos es objetivamente mejor a la basada en la indización *full-text*, en los siguientes términos:

- Los metadatos deben proporcionar instrumentos para describir el contenido semántico de un recurso y están mejor preparados para soportar la recuperación de información, que el propio documento (DLO). En muchos casos, los recursos de información no son capaces de facilitar en sí mismos, sus propias relaciones semánticas (por ejemplo, un código de un programa no puede facilitarnos cuál es su utilidad si no dispone de una caracterización en la que se describa).
- Los metadatos facilitan la "catalogación" o descripción de recursos no-textuales e información binaria, como: audio, software, imágenes, vídeo, etc., que supone una diferencia importante con respecto a la indización

¹²⁰ Bipin C. Desai. Supporting Discovery... *Op. cit.*, p. 193.

automática, dirigida sólo al texto. El interés por la búsqueda en distintos medios y en distintos dominios científicos del material digitalizado, evoca expresamente el concepto de metadatos y la posibilidad de definir SRII basados en elementos esenciales que sean comunes para todos los dominios y disciplinas, así como para todos los tipos de información (verbigracia, la imagen¹²¹, sonido, vídeo, entendidos todos ellos como DLOs).

- El hecho de existir gran cantidad de recursos electrónicos de carácter no-textual, implica la necesidad de SRII basados en metadatos, por ello, se han implementado arquitecturas de metadatos en entornos de recuperación de información multimedia¹²² basados en la Web, que permiten la formulación de preguntas sobre el contenido visual de la Red.
- Una categoría especial de metadatos que hace que los sistemas que los utilizan conlleven una ventaja comparativa en la recuperación, frente a los que buscan sobre el texto completo, son los metadatos sobre recursos que no son fácilmente accesibles porque están ocultos detrás de CGIs u otros *scripts*, o que son accesibles sólo a través de protocolos específicos

¹²¹ Con la aparición del amplio uso del Web, la imagen ha recibido una atención destacable, incluso en el ámbito de la imagen en movimiento, con proyectos pioneros como VISION (*Video Indexing for Searching Over Networks*). David Green. Beyond Word and Image. Networking Moving Images: More than Just the "Movies" [documento HTML]. *D-Lib Magazine*, July/Agust 1997, vol. 3, nº 7/8. Disponible en: <http://www.dlib.org/dlib/july97/07green.html> (consultado el 18 de junio de 1998).

¹²² Sobre la necesidad de metadatos para el acceso a bases de datos de contenido visual distribuidas en Internet, se están realizando diversos estudios, tanto desde una perspectiva semántica que podría estar representada por el sector de la información sobre patrimonio, como desde una perspectiva tecnológica llevada a cabo por Departamentos de informática, *V. gr.* Wendy Chang and Aindong Zhang. Metadata for Distributed Visual Database Access [documento HTML]. En: *IEEE Meta-data Conference (2. 1997. Silver Spring)*. IEEE Computer Society, 1997, rev. 17 de febrero de 1998. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/azhang/azhang.html> (consultado el 18 de junio de 1998) o Jane Hunter, Liz Armstrong. *Op. cit.*, <http://www8.org/w8-papers/3c-hypermedia-video/comparison/comparison.html>

(Internet invisible¹²³). Estos metadatos deben publicarse de tal forma que los programas que indizan sobre ellos puedan acceder fácilmente a la metainformación.

En sistemas automatizados de representación de información —Internet es, con todas sus particularidades, uno de ellos— la tradición bibliotecaria demuestra que la estructuración de la información y la búsqueda por campos (ya sea en OPACs o en bases de datos bibliográficas comerciales) facilita la búsqueda al usuario final y permite discriminar, aspectos de la semántica funcional del usuario como si se necesitan obras de un determinado autor o sobre un determinado autor, por ejemplo. Antonio Hernández define esta forma de búsqueda como *búsqueda cualificada*, y añade además que está *incomprensiblemente muy poco implementada y muy poco utilizada por los usuarios en aquellos motores, como Altavista, en que se permite*¹²⁴.

En el caso de los SRII basados en metadatos, la finalidad es la misma: facilitar un contexto semántico establecido *a priori*, de tal forma, que se puedan realizar búsquedas acotadas a distintos campos. Este es por ejemplo el caso, entre otros, de HotMeta <<http://www.dstc.edu.au/Research/Projects/hotmeta/search.html>> un SRII basado en metadatos, desarrollado de forma experimental por la Universidad de

¹²³ Todos los documentos pertenecientes a bases de datos diversas constituyen casi exclusivamente hoy en día, lo que denominamos Internet invisible. (Cfr. Capítulo 4, nota 178). La mayoría de los buscadores (Google, Wisenut, etc.) indizan ya documentos ofimáticos (.ppt, .doc, .xls, etc.) e información textual binaria (p. ej., documentos PDF o realizados con MS Word), sin embargo, existe todavía una gran cantidad de información, procedente de esas bases de datos accesibles vía Web, cuya información no es visible, ni recuperable a través de SRII de carácter global. Se estima que esa *Internet invisible* es 500 veces más grande que el total de la información indizada por los buscadores. Vid. <http://www.internetinvisible.com/quees.htm>

¹²⁴ Antonio Hernández Pérez. *Op. cit.*, p. 231. (La puntuación es nuestra). En la búsqueda cualificada de Altavista, no obstante, salvo en el caso de la posibilidad de buscar por título (title:), la calificación de las búsquedas se funda más en las posibilidades de búsqueda basadas en los DNS de los recursos (url:, domain:, host:) que en las supuestas metaetiquetas de las páginas que, como vimos (Cfr. *supr.* Nota 71) se excluyen expresamente para el cálculo de la relevancia de búsquedas genéricas.

Queensland (Figs. 56 y 57)¹²⁵. HotMeta basa su funcionamiento en un *search engine* que indiza según el Dublin Core. Al igual que otros sistemas basados en metadatos, permite acotar la búsqueda a un campo o elemento de la descripción, lo que da lugar a resultados más precisos.



Fig. 56. Interfaz de HotMeta (1998)

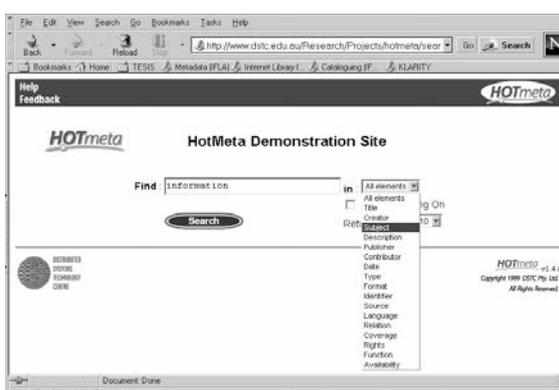


Fig. 57. Interfaz de HotMeta (1999-2001, v1.4.6.)

Los sistemas globales de recuperación de información en Internet basados en metadatos, comportan también riesgos:

- Supeditan en cierta manera la amplitud en las búsquedas a la calidad de las mismas, lo cual podría entenderse, en cierto sentido, como que los sistemas globales de búsqueda basados en metadatos adolecen de exhaustividad¹²⁶.

¹²⁵ Hemos elegido como ejemplo, un buscador que, a pesar de su carácter experimental o prototípico (proyecto universitario), tiene el propósito general de indizar DLOs descritos según el modelo DC. El *search engine* propiamente dicho, se utiliza para fundamentar la recuperación en algunas de las *subject gateways* australianas, V. gr. ADT (*Australian Digital Theses*) <<http://adt.caul.edu.au/combinedsearch.html>>. Vid. 8.2.2.

¹²⁶ Si tenemos en cuenta la terminología tradicional de la recuperación de información, comparar sistemas basados en metadatos *versus* sistemas basados en el texto completo, se puede asimilar a la comparación tradicional de lenguaje *natural versus lenguaje* controlado, o a la confrontación precisión *versus* exhaustividad, técnicas o ratios que son inversamente proporcionales en cada uno de los sistemas. El hecho de señalar que los SRI basados en metadatos son más precisos y menos exhaustivos como un riesgo aparente, se debe a que en contextos de información tan amplios como el de Internet, conseguir una alta razón de exhaustividad puede considerarse importante, aunque por otra parte es muy difícil saber realmente, como comentábamos al hablar de la relevancia (Cfr. *supr.* 7.1.2.), la tasa de exhaustividad en Internet.

- Por otra parte, los SRII basados en metadatos están diseñados por catalogadores o profesionales de la información y la mayoría de los elementos incluidos en ellos, aunque son útiles en términos generales, pueden resultar complejos y superar incluso el nivel de comprensión de los proveedores de información y del usuario final.
- En muchas ocasiones, la búsqueda basada en metadatos no es todo lo relevante que cabría esperar de un trabajo serio de análisis implícito del objeto de información ya que, a veces los propios autores/creadores de las páginas Web utilizan las metaetiquetas para el *spamming*¹²⁷. Este ardid se utiliza en algunos sectores comerciales para aumentar la visibilidad de sus sitios y atraer visitantes a sus páginas a través de falsas descripciones y redundancia de términos atractivos para la búsqueda en las etiquetas <META>. Este es uno de los grandes riesgos conlleva institucionalización de una búsqueda global basada en metaetiquetas, que ha llevado incluso a algunos autores, a decir que *el spamming es la maldición de los motores de búsqueda en la WWW*¹²⁸ y por ello, aunque se indexen sus contenidos, cada vez se utilizan menos para establecer el ranking de relevancia de los resultados.
- Recordemos además algo que apuntamos con relación a la creación de metadatos, donde algunos autores se pronunciaban en la tendencia a la dispersión que puede llegar a producirse cuando la cantidad de fuentes de

¹²⁷ Spam [documento HTML]. En: *Clear Presence: Web Promotion Workshop*. San Luis Obispo, CA: The Grid, rev. 9 de enero de 1998. Disponible en: <http://www.thegrid.net/clear/spam.htm> (consultado el 6 de agosto de 1998). En este capítulo tratamos el *spamming* como una amenaza genérica en los SRII que basasen sus búsquedas en metaetiquetas; asimismo hemos comentado, como la mayoría de los *search engines* con los que trabajan los principales motores de búsqueda, no tienen en cuenta las etiquetas. Cfr. 4.3.1. donde se trataba este problema desde el punto de vista de la creación de *metatags*. Para una definición exhaustiva de este concepto, Vid. Glosario, *spam*, *spamming*.

¹²⁸ *Spamming is the bane of WWW search engines*. Jessica Milstead and Susan Feldman. *Op. cit.*, p. 31.

información a las que se asignan metadatos es muy numerosa. En este caso, en las búsquedas en metadatos puede tener tan poco valor como las basadas en el texto completo¹²⁹.

- Finalmente, como reconocen varios autores¹³⁰, la creación de metadatos implica una inversión de tiempo y dinero. Los motores de búsqueda global de la WWW no los tienen en cuenta para incrementar la relevancia de sus búsquedas y por otra parte, los AU no necesitan esa metainformación para visualizar los contenidos, con lo cual, existen pocos incentivos para generar metadatos de autor. Así ha fracasado el objetivo inicial de metadatos de propósito general como el Dublin Core o las propias metaetiquetas básicas de HTML debido, en un principio, a la situación de *chicken-and-egg* que señalaba Gill y que citábamos en el capítulo anterior¹³¹ y posteriormente, a la utilización inconsistente o abusiva de esa metainformación básica.

A pesar de la utilidad de los sistemas de recuperación de información basados en una indización del texto completo, es necesario, cada vez más ante el crecimiento de la Red, el desarrollo de SRII de calidad, basados en metadatos. No obstante, basándonos en Taylor y Clemson¹³² podemos resumir las ventajas e inconvenientes

¹²⁹ Diego R. López y Javier Massa. *Op. cit.*, <http://www.rediris.es/rediris/boletin/45/enfoque1.html> Cfr. 4.2.

¹³⁰ V.gr. Tony Gill. Metadata and the World Wide Web [documento PDF]. En: *Introduction to metadata...* *Op. cit.*, <http://www.getty.edu/research/institute/standards/intrometadata/pdf/gill.pdf>, p. 16. Jean Hudgins, Grace Agnew, Elizabeth Brown. *Op. cit.*, p. 60.

¹³¹ Cfr. Capítulo 6, notas 70 y 71.

¹³² Arlene G. Taylor, Patrice Clemson. Access to Networked Documents: Catalogs? Search Engines? Both? [documento HTML]. En: *OCLC Internet Cataloging Colloquium (1996. San Antonio)*. OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/taylor.htm> (consultado el 19 de abril de 2001). Recordemos que este Coloquio sobre la Catalogación de recursos Web celebrado en el seno de OCLC, que ya hemos mencionado en otras partes de este trabajo, se celebró en el año 96, con lo cual, hemos tenido que modificar algunas de las características que señalan estas autoras porque en el transcurso de los últimos cinco años la realidad de estos SRII ha cambiado sustancialmente, aunque en esencia, su funcionamiento de indización del texto completo sea el mismo. Sobre los inconvenientes de motores de búsqueda globales, Cfr. *supr.* Nota 107, Abadal y Estivil.

que tienen los motores de búsqueda que indizan el texto completo. Dentro de los puntos fuertes, señalaban: que son gratuitos, que se puede obtener información sin procesamiento previo por parte del usuario; que permiten acceder a sus índices veinticuatro horas al día, siete días a la semana; que a pesar del dinamismo de Internet, a través de estos sistemas y de la potencia de sus robots se puede acceder a páginas recién creadas. También señalan como una ventaja, que los documentos que recuperan estos SRII, se encuentran inmediatamente disponibles, sin embargo, esto no es del todo cierto debido a la inconsistencia de los recursos Web y a que estos sistemas no siempre tienen un control actualizado sobre la mutabilidad de los DLOs. Por otra parte, resaltaban como puntos débiles la falta del control de vocabulario, que los recursos de búsqueda no muestran los contenidos de sus bases de datos ni anticipan los criterios utilizados para la indización, presentan dificultades para la recuperación de información por nombres de persona y entidad, a lo que hay que añadir, la dificultad para localizar información no-textual.

Podemos concluir que el problema de RI que subyace en la WWW, se ha provocado, entre otras diversas causas, porque se ha avanzado más en la potencia tecnológica de las aplicaciones, que en estructuración semántica de la información. El mayor reto ahora es desarrollar métodos que identifiquen y recuperen de manera unívoca y consistente la información de la Red, independientemente del formato que tengan y de cuál sea su ubicación. No obstante, los estándares de metadatos están en desarrollo y la Web Semántica, no es todavía más que la quimera hacia la que tiende la recuperación de información de carácter global en la WWW. Las iniciativas y normas tanto de estructuración de la información (XML/RDF) como de la semántica (DC), suponen un avance para arrostrar el desafío de una recuperación de información precisa y eficaz. Sin embargo el debate profesional que se ha generado sobre este tema, amén de la inestabilidad y complejidad de estos metadatos, nos hace pensar que la solución del problema de la recuperación cualificada, genérica y global sobre todos los recursos de la Red, dista mucho de ser una realidad a corto plazo. En cualquier caso, la investigación en la aplicación y normalización de metadatos, puede aportar un valor añadido a los SRII, sin olvidarnos del papel que juegan las

tecnologías *per se*, de cuya mejora, junto con el desarrollo de estándares, puede depender el éxito o no de sistemas de recuperación basados en metadatos¹³³.

Ante la realidad de la recuperación de información en Internet hoy en día podemos decir que los motores de búsqueda sobre el texto completo tienen una utilidad genérica, donde los metadatos (metaetiquetas) son simplemente un factor, más o menos contemplado para la indización. Los metadatos deben ser, sin embargo, la piedra angular de la organización de recursos electrónicos en entornos finitos, abarcables o bibliotecas digitales, que trataremos en el capítulo siguiente. Para ello, nos parece oportuno concluir este capítulo con dos de las proposiciones (P_1 y P_2) que fundamentan la RI, en la bibliografía tradicional, recogidas por Cortez¹³⁴, que adaptamos a continuación y que nos servirán para avalar la aplicabilidad de los metadatos en bibliotecas digitales o contextos finitos-locales-abarcables de información:

- P_1 : *La búsqueda de información será probablemente más eficiente y eficaz cuando el usuario esté familiarizado con la clasificación, estructura, contenido y propósito de la información buscada.*

¹³³ En este aspecto del papel de la tecnología (tendencia algorítmica en la recuperación de información) *versus* el papel documental (la tendencia cualitativa o cognitiva de la recuperación de información), existen diversas posturas: el "tecno-escepticismo", y que podría estar representada por Kuny y Cleveland, o incluso Oddy, que piensan que el desarrollo tecnológico *juega sólo un pequeño papel* (Kuny y Cleveland. *The Digital Library: Myths and Challenges*. *IFLA Journal*, 1998, vol. 24, nº 2, p. 109.) y otra tendencia que podríamos llamar "tecno-*praesentia*" representada por aquel sector que confía sólo a las máquinas la posibilidad de la recuperación, sin embargo, pensamos que la solución óptima estará representada por una conjunción de ambas tendencias: normalización bibliotecaria + desarrollo de la tecnología, uniéndonos con esto al pensamiento de Lynch: *Combining the skills of the librarian and the computer scientist may help organize the anarchy of the Internet*. Clifford Lynch. *Searching...* *Op. cit.*, <http://sciam.com/0397/issue/0397lynch.html>

¹³⁴ Edwin M. Cortez. Use of Metadata Vocabularies in Data Retrieval. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, nº 13, p. 1218, 1223. Para la primera proposición (P_1) se basa en autores como Belkin, R.N. Oddy & Brooks, Borgman y Kuhlthau, mientras que para la segunda (P_2), lo hace en Bates, Harter y Svenonius.

- P_2 : El proceso de recuperación de información conlleva la definición (el establecimiento de los límites de) un conjunto de documentos de una colección, q e tiene un límite definido.

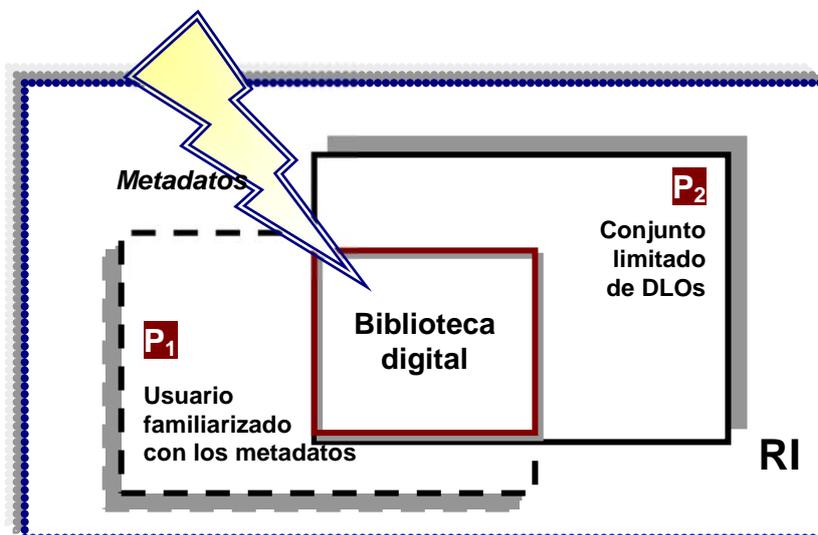


Fig. 58. Visión de la biblioteca digital según las proposiciones clásicas de la RI

CAPÍTULO 8:

BIBLIOTECAS DIGITALES Y METADATOS

La World-Wide Web ha cambiado la forma de dar servicio en las bibliotecas. Para mantener sus funciones principales y generar servicios en el nuevo entorno de red, las bibliotecas han desarrollado de forma vertiginosa sitios Web, han colocado sus OPACs en la Web (WebPACs), han establecido importantes servicios de información desde el punto de vista cuantitativo y también, como señalábamos en el capítulo anterior, han creado sistemas de recuperación de información desde el punto de vista cualitativo. Al mismo tiempo, el crecimiento de la Web ha proliferado tan rápidamente y en tantos aspectos de la sociedad que distintos sectores profesionales reconocen la necesidad urgente de organizarla. Debido a ese rápido crecimiento los bibliotecarios y otros profesionales de la información están desarrollando diversas soluciones para controlar la explosión de recursos Web. Para trabajar en este nuevo contexto, necesitan aprehender los principios generales en lo relativo a la creación de contenidos digitales y metadatos asociados a ellos, para poder desarrollar nuevos sistemas de recuperación de información y participar en un desarrollo de sistemas de información "de calidad". Basándonos en estas circunstancias, a lo largo de este capítulo trataremos de caracterizar las bibliotecas digitales en función de los metadatos y del papel que estos desempeñarán en el cambio de paradigma en la gestión bibliotecaria.

Como noción clave en la redefinición de las bibliotecas en la Sociedad de la Información, el concepto de biblioteca digital está tomando cada vez más, un papel fundamental en la planificación estratégica de nuevos servicios de información, a la par que surgen bibliotecas exclusivamente digitales con la finalidad de facilitar el acceso universal al patrimonio científico y cultural. A pesar de la diversidad y

heterogeneidad de realidades informativas en la Red llamadas "biblioteca digital, virtual o electrónica", en nuestra tesis, partimos de cuatro ideas básicas para precisar este concepto, en virtud de los metadatos:

1. Aunque son muchos los autores que de forma intuitiva y quizás, a nuestro juicio, promocional, definen la WWW como una biblioteca digital¹, Internet, la WWW, en el planteamiento de esta tesis, no es, ni puede identificarse con "LA" biblioteca digital.

Varios autores, como Lynch² o Koehler³ ratifican nuestra afirmación. El símil de Internet como una biblioteca mundial en la Era digital no debe prevalecer ni siquiera en un examen superficial de la Red. Internet, la WWW, no fue diseñada para la publicación y recuperación organizada de información como hacen las bibliotecas tradicionales. Ha evolucionado —dice Lynch⁴— *hacia una idea de archivo documental caótico para la producción colectiva de la imprenta digital mundial*. Ese depósito de información contiene no sólo libros y artículos, sino datos científicos originales, grabaciones de audio, vídeo, conversaciones interactivas, etc., material efímero mezclado por todo el

¹ Por ejemplo, Lesk dice expresamente que la Web es la biblioteca digital mundial (*We now have a world-wide digital library, the Web*). Michael E. Lesk. The Organization of Digital Libraries. En: *Digital Libraries: Philosophies... Op. cit.*, p. 10.

Brisson justifica afirmaciones como la de Lesk, en su caso, citando a Michael Hart del Proyecto Gutenberg <<http://www.promo.net/pg>> argumentando que, por su naturaleza "virtual" muchos profesionales han caído en la tentación de igualar la World Wide Web en sí misma a una biblioteca digital gigante. Roger Brisson. *Op. cit.*, p. 5.

² Clifford Lynch. Searching... *Op. cit.*, <http://sciam.com/0397/issue/0397lynch.html>

³ Wallace Koehler. Digital Libraries and World Wide Web Sites Page Persistence [documento HTML]. *Information Research*, June 1999, vol. 4, n° 4. Disponible en: <http://www.shef.ac.uk/~is/publications/infres/paper60.html> (consultado el 29 de septiembre de 1999). Koehler matiza acertadamente, que la Web no es una biblioteca, pero sí se puede seleccionar de la Web material para las colecciones de las bibliotecas. Este principio, la Web como foco de selección de DLOs se sigue en la formación de *subject gateways* y de bibliotecas híbridas. *Vid. infr.*

⁴ Clifford Lynch. Searching... *Op. cit.*, <http://sciam.com/0397/issue/0397lynch.html>

mundo con trabajos importantes desde el punto de vista de su permanencia y su calidad científica.

2. Para poder darle el calificativo de "biblioteca digital" a un sitio, sede o portal Web en Internet, tienen que darse las condiciones que justifiquen tanto el sustantivo (biblioteca) como el determinante (digital), en estos términos⁵:
 - Biblioteca: debe implicar, selección, organización y colección.
 - Digital: debe implicar que los ítems de la colección estén en formato de codificación numérica, y que la organización y acceso a todos los servicios sea también digital, a través de la Web.
3. Si una biblioteca es un conjunto *organizado* de documentos, una biblioteca digital debe ser un conjunto *organizado* de documentos electrónicos, o mejor dicho, un conjunto organizado de DLOs, donde los metadatos serán la clave de esa organización en aras a mejorar el acceso a la información. Lo que diferencia a una biblioteca de una librería o conjunto de libros es justamente la organización y control de los documentos que la constituyen. De forma análoga, lo que debe diferenciar una biblioteca digital de una página/sitio/portal Web⁶, de un conjunto de enlaces más o menos organizados, o de un sistema de búsqueda sobre el texto completo —que serían las tres posibilidades de supuesta organización y/o disposición convencional de recursos electrónicos— debe ser justamente su organización

⁵ Las dos condiciones que señalamos podrían caracterizar perfectamente sistemas de recuperación de información como INFOMINE o BUBL, como bibliotecas digitales. De hecho, como apuntamos en el capítulo anterior son muy escasas las fronteras entre estos servicios de información digital y lo que entendemos aquí por bibliotecas digitales. Apenas su cobertura enciclopédica y no limitada nos hace aproximarlos más a los SRII genéricos, si bien, el carácter bibliotecario y la concepción del servicio de información nos hacen considerarlos SRII "de calidad".

⁶ Entendemos también dentro de sitio Web todos los portales que, de forma genérica (portales horizontales), o temática (vortales), pretenden ser el punto de acceso a los contenidos y a los servicios de la WWW. *Cfr. infr.* Concepto de portales bibliográficos o bibliotecas digitales como "servicios de valor añadido".

y control a través de metainformación, sea ésta dependiente del recurso o del contenido.

4. Las bibliotecas digitales, tal y como las entendemos aquí, son un contexto abarcable y operativo de recursos de información electrónicos organizados, donde la aplicación de un sistema de metadatos sirve para organizar esos DLOs, contribuyendo a mejorar la recuperación y el intercambio de información. Los metadatos son pues, en este contexto finito, el sustento de sistemas informativos digitales de calidad o *el elemento que permite que grandes colecciones de documentos y datos funcionen como bibliotecas organizadas en vez de desorganizados depósitos de chatarra*⁷.

Partiendo de estas cuatro afirmaciones, dedicaremos este capítulo a abordar la problemática vinculada a las bibliotecas digitales tanto desde el punto de vista del concepto y de su organización, como de sus diferentes tipos y proyección de futuro a tenor de la Sociedad de la Información y del acceso global a los contenidos. Si bien es cierto que las bibliotecas digitales podrían estudiarse desde innumerables puntos de vista, y que cada uno de ellos darían lugar a múltiples investigaciones de índole semejante a esta, abordaremos el tema desde la visión de la nueva organización de la información electrónica y de la aplicabilidad de los metadatos ante el cambio de paradigma en la gestión de información digital. Así, concebimos un aval teórico para defender el proyecto de aplicación de metadatos a una biblioteca virtual que proponemos en el capítulo 9 de esta tesis doctoral.

⁷ *Meta data is the essential "glue" which enables large collections of documents and data to function as organized libraries rather than disorganized junkyards.* Kal Ahmed, et al. *Op. cit.*, p. 11. Transcribimos esta elocuente afirmación sobre los metadatos por un lado porque respalda una de las hipótesis más importantes de nuestra investigación (el carácter fundamentante de los metadatos en el ámbito de las bibliotecas digitales y en el cambio de paradigma en la gestión de la información al que venimos aludiendo en toda esta tesis) y por otro porque es sorprendente, a la par que gratificante, que investigadores del mundo de la informática (Kal Ahmed, Danny Ayers, Mark Birbeck, Jay Cousins, David Dodds, Josh Lubell, Misoslav Nic, Daniel Rivers-Moore, Andrew Wat, Robert Worden y Ann Wrightson, coautores de la obra citada) coincidan en nuestra visión de los metadatos en el contexto de las bibliotecas digitales.

8.1. ¿Biblioteca digital/virtual? La biblioteca como concepto y la diversidad designativa

A lo largo del tiempo, el papel, el material impreso desde Gutenberg, ha sido el soporte físico substancial de los fondos y servicios de las bibliotecas tradicionales. Esas bibliotecas basadas en papel se han ido automatizando gradualmente, dando lugar a una realidad en la que la colección está basada en átomos de información (fundamentalmente libros y papel), pero los servicios están automatizados (catálogos de acceso público en línea, préstamo automatizado, etc.). En poco tiempo Internet ha potenciado aún más el nivel de automatización de servicios y procesos en las bibliotecas (desideratas a través de formularios Web, catalogación por copia a través de Z39.50, etc.). Pero a raíz de Internet, no sólo los servicios, también la propia información, las colecciones, se crean, se procesan y se difunden en forma electrónica, a través de las redes. Toda esa información en bits, esos grandes almacenes de datos e información se han empezado a llamar biblioteca digital, virtual o electrónica dando lugar a una complejidad y diversidad designativa. En toda esa nueva terminología que trata de nominar diversas realidades informativas basadas en la Web entendemos la biblioteca como concepto.

El entorno cambiante provocado por las tecnologías de la información y de las comunicaciones (TICs), está mutando paulatinamente la visión de la biblioteca como sistema de almacenamiento y recuperación de información, hacia una visión más integradora e interactiva, como un sistema de comunicación. Las funciones de las bibliotecas se están transformando a la vez que confían la Web como sistema de distribución de información. El progreso de la tecnología ha cambiado el *cómo* los bibliotecarios hacemos nuestro trabajo, no el *porqué*. Kuny y Cleveland⁸ piensan que la tecnología no alterará substancialmente el trabajo de los bibliotecarios y

⁸ Terry Kuny and Gary Cleveland. The Digital Library: Myths and Challenges. *IFLA Journal*, 1998, vol. 24, nº 2, p. 107.

documentalistas: *relacionar a la gente con la información*, lo que sí cambiará y ha cambiado ya, es la forma de hacerlo. A pesar de esta opinión filantrópica, que en cierta medida subyace a la filosofía bibliotecaria tradicional, según nuestro punto de vista, el desarrollo tecnológico más profundo, la interconexión mundial de ordenadores, puede alterar el concepto fundamental de la biblioteca en nuestro nuevo siglo XXI. Quizás haya que replantear no sólo el cómo sino también el cuándo, el dónde, etc. del trabajo del profesional "documentalista-bibliotecario-gestor de información" (de información electrónica).

Los bibliotecarios han organizado el conocimiento gracias al proceso del análisis de contenido y de la descripción bibliográfica, obteniendo como resultado registros de información sobre la información, en definitiva, metadatos en sentido amplio como apuntábamos en el capítulo 3. Como señala Oddy⁹, *todo sistema de acceso bibliográfico se caracteriza por tener dos elementos esenciales: 1) una infraestructura técnica y una interfaz de usuario y 2) datos bibliográficos estructurados* (= metadatos). En esta misma línea, Ellis y Ford¹⁰ precisan que, para una recuperación eficiente, eficaz y sencilla de documentos útiles, se ha determinado tradicionalmente la construcción de sistemas compuestos por los siguientes elementos estructurales:

- Una base de datos constituida a partir de registros, que corresponden a la representación de un documento (o a una parte de un documento) de la colección.
- Un mecanismo comúnmente conocido como motor de búsqueda (*search engine*) que identifica las relaciones existentes entre cada registro de la base de datos y la pregunta del usuario, y que ordena los resultados según tal comparación.

⁹ Pat Oddy. *Op. cit.*, p. 154.

¹⁰ David Ellis and Nigel Ford. *Op. cit.*, p. 28.

- Una interfaz que permite al usuario controlar los procesos de formulación y lanzamiento de la consulta y la presentación de los resultados.

Estos elementos constitutivos de un sistema de acceso a la información, así como el objetivo tradicional de organización y recuperación de información no han cambiado esencialmente, e incluso, como señalábamos en el capítulo anterior, han servido de base en la arquitectura funcional de muchos de los SRII. Sin embargo, a tenor de esos sistemas de recuperación de información en Internet, se ha dedicado más esfuerzo a mejorar las interfaces y a desarrollar potentes motores de búsqueda, que al análisis y representación de la información.

Desde los inicios de la popularización de la WWW como entorno informativo, las bibliotecas han mantenido catálogos de acceso público Web por un lado, y listas electrónicas de recursos (enlaces a revistas electrónicas, bases de datos, y otros recursos de interés para la comunidad a la que sirven) por otro. La calidad de los recursos electrónicos disponibles en la Web está teniendo un impacto directo en la formación de las colecciones bibliotecarias. En algunos casos, esas "recopilaciones de enlaces" iniciales han dado lugar a verdaderos servicios de información que se han gestado sobre la base del criterio selectivo bibliotecario y sobre el compromiso de dar acceso cualificado a los recursos de información de Internet. Muchas veces inclusive, esa pretensión bibliotecaria (normalmente vinculada a bibliotecas universitarias) ha crecido de tal forma que ha originado servicios como BUBL o INFOMINE que sirven de punto de referencia y de sistemas de recuperación genéricos de Internet¹¹. En otras ocasiones las bibliotecas tradicionales han confiado la recuperación de información en Internet a buscadores, directorios y otros de los "grandes" SRII ya convencionales, al incluirlos normalmente entre sus enlaces¹². Salvo algunas

¹¹ *Cfr. sup.* Nota 5.

¹² Sin embargo, como ya mencionamos en el capítulo anterior, el funcionamiento y el rendimiento de estos buscadores es insatisfactorio desde un punto de vista estrictamente bibliotecario. *Cfr.* Ernest Abadal, Assumció Estivil. *Op. cit.*, <http://www.cbuc.es/5digital/ireweb.pdf>

excepciones en que, sobre la base de proyectos como CORC, se integran registros de metadatos en catálogos MARC, no se ha asumido de forma generalizada el reto de agrupar ambos tipos de información en un sólo servicio híbrido e integral.

De forma creciente, las bibliotecas incluyen también en sus sitios Web diversos recursos electrónicos tanto comerciales (fundamentalmente, bases de datos y revistas electrónicas), como gratuitos (publicaciones periódicas, actas de congresos, información relativa a instituciones señeras en un ámbito científico, etc.). Toda esta prodigalidad informativa en torno a la Red, junto a la pretensión bibliotecaria de proporcionar un acceso autorizado también a sus contenidos, ha promovido el desarrollo de diversas realidades que, cada vez más, se denominan biblioteca electrónica, digital e incluso, como ya hemos apuntado, virtual. Además, la facilidad de digitalización ha hecho que cualquier institución pueda disponer de una supuesta biblioteca digital formada por recursos de índole diversa digitalizados por distintos motivos¹³. Estas experiencias tienen en algunos casos un carácter estratégico y en otros, suponen simplemente soluciones aisladas de gestión de la información electrónica difundida a través de la WWW.

Según Smith¹⁴, tal y como se conciben normalmente, *las bibliotecas digitales son bibliotecas en las que las colecciones están en forma digital y el acceso a la información contenida en ellas está basado totalmente en la tecnología digital*, a lo que nosotros matizamos que la organización y descripción de los datos debe producirse originalmente también de forma digital, a través de metadatos. La noción

¹³ Por ejemplo, en la Organización de Estados Iberoamericanos (OEI) se denomina "biblioteca digital" a un conjunto de publicaciones digitalizadas editadas originalmente en papel por la propia institución y que por diversas razones se encuentran agotadas o son difíciles de localizar. Esta "biblioteca digital" está constituida por una serie de enlaces a dichas publicaciones que previamente han sido escaneadas y convertidas a formato .pdf. A lo sumo, se puede buscar entre ellas por palabra clave a través de la tecnología Googlebot adoptada como sistema de recuperación en esta biblioteca. Vid. <http://www.campus-oei.org/oeivirt> y <http://www.campus-oei.org/oeivirt/buscador.htm>

¹⁴ Terence R. Smith. *Op. cit.*, <http://www.dlib.org/dlib/july96/new/07smith.html>, Cfr. Capítulo 3, nota 49.

clave en la redefinición de las bibliotecas, y de otros aspectos que caracterizan la contemporaneidad en la Sociedad de la Información, es pues "digital".

Conforme al DRAE¹⁵, digital, en su segunda acepción, significa: *referente a los números dígitos y en particular a los instrumentos de medida que la expresan con ellos*. El significado en nuestro idioma de este término, no nos da muchos vestigios de lo que puede conllevar la tecnología digital, y mucho menos si tomamos el ejemplo que da dicha obra de referencia: *p. ej. reloj digital*. Tenemos que recurrir a un diccionario especializado¹⁶ para aclarar el concepto en el ámbito de las TICs, donde digital se refiere a *la nueva tecnología que genera, almacena y procesa datos en dos estados: positivo y no-positivo*; de tal forma que los datos transmitidos o almacenados según esta tecnología se expresan en cadenas de unos y ceros que conforman los bits, la unidad mínima de información para un ordenador. Para entender las dimensiones de esta nueva tecnología contraponemos la comunicación digital a la analógica, de la misma forma que enfrentamos bits y átomos, en aras a marcar un punto de referencia que nos permita confrontar la Era Gutenberg con la Era Web:

- En lo que, de forma genérica, podemos denominar Era Gutenberg, la información estaba constituida por átomos en forma de documentos, se organizaba en bibliotecas y se daba acceso a través de catálogos (estos últimos normalmente automatizados).
- Mientras que, en lo que ya denominamos Era Web, entendemos que la información está formada por bits en forma de DLOs, se debe organizar en bibliotecas digitales y dar acceso a través de metadatos, que también son bits.

¹⁵ *Diccionario de la lengua española* [documento HTML]. 22ª ed. Madrid: Real Academia de la Lengua, 8 de octubre del 2001. Disponible en: <http://buscon.rae.es/drae/drae.htm> (consultado el 16 de octubre de 2001).

¹⁶ Whatis?com: http://whatis.techtarget.com/definicion/0,,sid9_gci211948,00.html (consultado el 16 de octubre de 2001).

La tecnología digital ha cambiado las características organizacionales de la sociedad, y sobre todo de las comunicaciones. El entorno digital proporciona, como reconocen Terceiro y Matías¹⁷, tres tipos de avances iniciales: *manejar de forma única toda clase de fuentes de información; una mayor flexibilidad y libertad en la incorporación de nuevos servicios; y más efectividad para almacenar y procesar información*. Estas tres mejoras que supone el digitalismo, implican una oportunidad categórica, al menos en principio, para la gestión de la información ya que en ellas se presume por un lado, una gestión y acceso uniformes a diversos recursos informativos, por otro, una expansión de la principal función de la biblioteca, dar servicio. Así, la tecnología digital ha cambiado también las características organizacionales de las bibliotecas tradicionales al menos en cuatro aspectos básicos:

- a) Por un lado, la organización de la información física: en lugar de realizar una organización material de los DLOs, se realiza una organización lógica más flexible.
- b) Por otro, esta organización de la información es múltiple, ya que un documento entendido como objeto de información, se puede almacenar según distintos criterios organizacionales, también lógicos.
- c) Pero sin duda alguna el cambio más significativo se produce en el ámbito de la metainformación y de la política de descripción de DLOs¹⁸ de la biblioteca. El hecho de que los documentos entendidos como objetos de información estén en formato digital permite la extracción directa de los metadatos, o bien el almacenamiento conjunto de estos con los datos. Esa información extraída puede satisfacer directamente las necesidades de los usuarios o puede utilizarse para la

¹⁷ José B. Terceiro y Gustavo Matías. *Op. cit.*, p. 81.

¹⁸ Seguimos manteniendo la convicción, desarrollada a lo largo de todo este trabajo, de que los documentos de la Red deben entenderse como objetos de información.

selección y caracterización de una colección de DLOs, como por ejemplo, para formar *subject gateways* o bibliotecas virtuales temáticas¹⁹.

d) Por otra parte, la automatización del acceso a la información lleva como contrapartida la pérdida de interacción directa entre el usuario y el gestor de la información o bibliotecario²⁰.

Capurro²¹ ve en las nuevas bibliotecas digitales un valor añadido con respecto a las bibliotecas tradicionales, incluso aludiendo como ventajas lo que habíamos citado como inconveniente. Sistematiza esta visión *del valor agregado* de estos nuevos servicios de información de la siguiente manera:

- *Acceso a documentos independientemente del espacio y el tiempo*²².
- *Combinación de documentos (textos de diversos tipos, incluyendo documentos multimediales) en diversos archivos digitales y clásicos.*
- *Búsqueda de documentos e informaciones no digitalizadas, utilizando catálogos en línea de acceso público (ahora WebPACs²³).*

¹⁹ Vid. *infr.* 8.2.2.

²⁰ Sin embargo, también se podría ver en esta característica de la información electrónica el efecto contrario al poder contactar el bibliotecario, a través del correo electrónico, formularios, etc. con un número infinito de usuarios potenciales de la biblioteca digital. En este sentido, la biblioteca Virtual *Miguel de Cervantes* que hemos elegido como objetivo para la aplicación de nuestra investigación en metadatos contempla esta posibilidad a través de un servicio específico de comunicación con el bibliotecario: <http://www.cervantesvirtual.com/bibliotecario> Vid. 9.2.1.

²¹ Rafael Capurro. *Las bibliotecas en la Era Digital* [documento HTML]. Capurro.de, rev. 6 de diciembre de 2000. Disponible en: <http://www.capurro.de/gohete.htm> (consultado el 29 de diciembre de 2000).

²² Berners Lee ratifica esta idea diciendo que *Internet y el Web nos han sacado del espacio bidimensional*. Tim Berners Lee. *Tejiendo...* *Op. cit.*, p. 186.

²³ Una visión muy interesante en torno a la evolución de los catálogos, entendidos como una de las estructuras básicas de comunicación en los servicios bibliotecarios del siglo XX la encontramos en: Michael Gorman. *From Card Catalogues...* *Op. cit.*, http://www.unicamp.br/bc/gorman_paper.html

- *Combinación de los procesos de información y comunicación: por ejemplo contactos a través de correo electrónico con el/los autor(es)/la(s) autora(s) de un documento y/o con el bibliotecario*²⁴.
- *Creación de grupos de intereses culturales, científicos, políticos etc. basados en técnicas digitales (newsgroups, mailing lists etc.).*

Quizás ese "valor agregado" o añadido que reconoce Capurro a las bibliotecas digitales sea lo que haya llevado a difundir, en algunos foros profesionales, su semejanza a los portales e incluso, el concepto de portal bibliográfico, donde se unen la idea de portal (=servicios de valor añadido a la búsqueda) y bibliográfico (=repertorio de libros, en este caso digitales, sobre una materia). El universo de información digital, potencia la función de servicio de las bibliotecas.

El concepto de portal está vinculado a la integración de servicios, a la reactivación de nuevos mecanismos de difusión, a la difusión selectiva de la información a través del concepto de personalización, a la creación de guías Web organizadas y al establecimiento de foros de comunidades virtuales de usuarios. Todas estas características pueden confluir en una biblioteca digital o virtual, y así lo reconocen algunos autores²⁵, e incluso, llevando al extremo la contemporaneidad terminológica, concebir —como Thomas²⁶— los *catálogos bibliotecarios como portales de*

²⁴ *Cfr. sup.* Nota 19.

²⁵ Hal Vergel A Digital Library as a Portal [documento HTML]. *DL Pearls*, June 2001. Disponible en: http://www1.acm.org/dl/dlpearls/dlp_6-01.html (consultado el 27 de julio de 2001).

Por ejemplo, en la reunión de usuarios de OCLC de febrero de 2001, se discutió la posibilidad de construir y mantener un portal bibliotecario a partir del producto de OCLC *WorldCat* (Vid. Capítulo 4, nota 34), ya que permite la integración completa de recursos, las referencias cooperativas y posiciona las bibliotecas en la Web. Patricia A. Wand. *Report from the OCLC Users Council, February 11 – 13, 2001* [documento PDF]. *CAPCON ALERT Special Edition*, March 2001, n° 65. <http://www.capcon.net/alertArchives/March%202001.pdf> (consultado el 24 de abril de 2001).

²⁶ Sarah E. Thomas. The catalog as portal to the Internet [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 4 de [cont.]

conocimiento. Esta concepción de portal asociado a las bibliotecas digitales está muy asumida en la realidad, aunque no siempre se utilice el término "portal" por estar éste excesivamente vinculado al nuevo modelo de negocio que constituyen los portales horizontales genéricos de Internet, y también, cada vez más, los vortales.

Las bibliotecas digitales son, no obstante, el contexto idóneo para desarrollar servicios de valor añadido dirigidos a una comunidad de usuarios local (institucional, por ejemplo, una comunidad universitaria concreta), o a una comunidad virtual temática distribuida (v.gr. las *subject gateways* que en cierto sentido podrían entenderse como portales bibliográficos²⁷). A ello se une también la amplitud del término portal y su uso, acerbamente extendido, para designar cualquier sitio o sede Web con cierta pretensión de dar servicio electrónico o simplemente información organizada. Las características de estos portales, según el proyecto británico Hylife²⁸, que aproximan dicho concepto de organización de la información más a las bibliotecas digitales (*subject gateways*) que a los buscadores genéricos, son:

- Más que intentar cubrir todos los recursos accesibles en la Web, un portal incluirá servicios seleccionados que han sido elegidos para llamar la atención de los usuarios a los que intenta servir el portal.
- Los recursos incluidos en un portal se ordenan en forma directorio, a los que se accede a través del *browsing* dirigido o navegación²⁹ a través de hiperenlaces, sin que ello obste para incluir también una interfaz de búsqueda por palabras clave.

septiembre de 2000, rev. 16 de noviembre de 2000. Disponible en: http://www.unicamp.br/bc/thomas_paper.html (consultado el 21 de febrero de 2001). Cfr. 9.2.1., donde se analiza el concepto de portal en la Biblioteca Virtual *Miguel de Cervantes*, frente al de comunidad virtual de usuarios.

²⁷ Así se ve por ejemplo en la documentación relativa al proyecto Hylife, donde expresamente se dice que *el concepto de portal Web puede verse como análogo al de subject-gateway*. Vid. <http://hylife.unn.ac.uk/toolkit/Portals.html>

²⁸ *Ibid.*

²⁹ *Vid.* 7.1.1.1.

- El portal, si el servicio así designado tiene alguna pretensión defendible relacionada con el término utilizado para denominarlo, indicará que actúa de cierta forma que garantiza la calidad y fiabilidad de los recursos a los que da acceso. Y este es el aspecto que aproxima a los portales a las bibliotecas digitales temáticas y distribuidas denominadas *subject gateways*³⁰.

En cualquier caso, la biblioteca en el contexto de Internet es un concepto, de la misma forma que la Web es un lugar, no es sólo forma, sino función. La biblioteca es única en la sociedad en tanto que no sólo selecciona, organiza, almacena y recupera información, también proporciona acceso, protege la libertad intelectual y proporciona asistencia directa e instrucción (e interacción) a sus usuarios³¹.

¿Qué es en realidad una biblioteca digital, virtual o electrónica?

Como decía Stephen P. Harter³² en 1997, *no está claro qué se quiere decir con el término biblioteca digital*. Esta situación no ha cambiado mucho en últimos años, se utiliza profusamente el término, pero pocas veces se define o caracteriza. Se sigue aplicando a una amplitud extraordinaria de aplicaciones, desde colecciones de revistas electrónicas, agentes de software, hasta a la página Web de una biblioteca tradicional, e incluso, como hemos mencionado, a toda la Web/Internet.

³⁰ Esta afirmación pone de manifiesto lo que anunciamos en el epígrafe de este apartado, "la diversidad designativa" que acontece no sólo a las bibliotecas digitales, sino también a la generalidad de los servicios de Internet, donde con bastante frecuencia se denomina portal a cualquier sitio o página Web, por ejemplo el portal de bibliotecas de la UNESCO *Cfr. infr.* Nota 30 y 8.2.2.

³¹ *The library is a concept as Web as place –it is function, not form alone. The library is unique in society in that it not only selects, organizes, stores and retrieves information, it also subsidizes access, protects intellectual freedom, and provides direct assistance and instructions to its users. The Academic Library in the Information Age.* [S. 1]: CETUS, 1997, p. 1. Citado por Lluís M^a Anglada i de Ferrer. Biblioteca Digital ¿mejor, peor o sólo distinto? *Anales de Documentación*, nº 3, 2000, p. 39.

³² Stephen P. Harter. *Op. cit.*, <http://jodi.ecs.soton.ac.uk/Articles/v01/i01/Harter>

Aunque no es objetivo de esta tesis hacer una reflexión profunda sobre el concepto de biblioteca digital/virtual/electrónica, sí es importante matizar algunos aspectos que contribuyan a esclarecer nuestra teoría de biblioteca digital como "conjunto organizado de DLOs". Para ello, en primer lugar, vamos a sistematizar y explicar los factores que a nuestro juicio, complican la determinación de lo que es una biblioteca digital:

- El propio término, "biblioteca virtual", es un oxímoron que dificulta su conceptualización. En la figura que se llama oxímoron, se aplica a una palabra un epíteto que parece contradecirla; así los gnósticos hablaron de una luz oscura; los alquimistas, de un sol negro... y los bibliotecarios, comprometidos con el reto de la Sociedad de la Información, hablamos de bibliotecas virtuales, donde se fusionan elementos en conflicto (biblioteca y virtualidad), desestabilizando el sentido e instalando una contradicción semántica.

Esta característica de la propia composición del término en forma de oxímoron dificulta la aprehensión del concepto. En muchas ocasiones, quizás por su naturaleza "virtual", se cae en la tentación de igualar la World Wide Web en sí misma con una, con la gran biblioteca digital; o también se tiende a denominar de forma muy espontánea "biblioteca digital/virtual" a cualquier servidor Web que albergue textos u otros materiales en formato electrónico.

- Otro de los factores que dificultan la definición de biblioteca digital/virtual, es el propio halo metafórico que acompaña al mundo de Internet como medio y recurso de información. Ese carácter metafórico que se utiliza para hablar de Internet, se ha potenciado además por las políticas públicas de acceso global a la información (*GIS, Global Information Society*) y por estudios de carácter filosófico-predictivo de principios de los 90's como *La aldea global* de Alvin Toffler, que lleva a equiparar la World Wide Web, con la biblioteca universal. Como ya hemos precisado, Internet no puede asimilarse a una "Gran Biblioteca Digital Mundial", salvo en símiles literarios y/o filosóficos que hablen de la Red como un mundo informativo global. En todo caso, siguiendo con los tropos que

refieren a Internet, denominaciones como esa (Gran Biblioteca Digital Mundial) no sería más que una bonita y sutil metonimia que toma el efecto (facilitar el acceso a la información, propio de una biblioteca) por la causa, la interconexión universal de redes que soportan información. No obstante, a pesar de ello es factible y deseable que las bibliotecas digitales interoperen de tal forma que se universalicen o globalicen más sus servicios³³.

- La diversidad de sitios o páginas Web que se autodenominan "biblioteca digital" u otros sinónimos alusivos, que evocan de alguna manera la palabra, y por ende el concepto, de biblioteca. Cualquier sitio que implique un conjunto de "libros", obras de la literatura, o simplemente recursos de información en la WWW reciben la denominación formal e intuitiva de biblioteca digital, cuando en el fondo no son más que un almacén de datos electrónicos o textos digitales, un conjunto temático de enlaces o, en el mejor de los casos un conjunto de textos o documentos pasados a formato digital y accesibles en una página HTML.

Verbigracia, el portal de la UNESCO sobre Bibliotecas³⁴ recoge siete sitios bajo el marbete de "bibliotecas digitales" en español. A modo de ejemplo analizamos cada uno de ellos para ilustrar las diferentes realidades que se asimilan:

- 1) Biblioteca VirtuArt <<http://biblioteca.8m.com/index1.htm>> Forma parte de los servicios que da un pseudo-portal valenciano denominado Hort@ Virtu@rt <<http://meliana.8m.com>>. Recoge tan sólo siete documentos (cuatro en

³³ Cfr. 8.3.

³⁴ UNESCO Libraries Portal <http://www.unesco.org/Webworld/portal_bib> Este sitio de la UNESCO se autodefine como "una pasarela de información internacional para bibliotecarios y usuarios de biblioteca" (*An international information gateway for librarians and library users*).

Las bibliotecas digitales a las que nos referimos en el ejemplo se encuentran en la secuencia jerárquica del directorio de este portal: Preservation and Access initiatives: Access: Digital libraries: Collections: Spanish. (No consignamos el URL completo por su excesiva longitud). Fecha de acceso: 24/10/2001. No respetamos el orden alfabético que se da en el portal a las supuestas "bibliotecas digitales".

castellano y tres en valenciano) de distinta índole a los que se presenta un enlace hipertextual dentro del mismo servidor. Desde ningún punto de vista podríamos considerar este sitio una biblioteca digital, sin embargo se denomina así, quizás por la ingenuidad o elementalidad de que da acceso a libros.

- 2) Ciberoteca de la obra social de Bancaja <<http://www.ciberoteca.com>>. Es un sitio que trata de emular los servicios básicos de una biblioteca, incluso dividiendo sus servicios con las denominaciones propias de las bibliotecas convencionales: sala de consulta, sala de reuniones y sala de lectura (Fig. 59). En cuanto al hecho de que se incluya bajo la denominación de "biblioteca digital" en ese portal de la UNESCO, podemos decir que en ella confluyen las dos realidades que contemplamos en esta tesis como biblioteca digital³⁵: por un lado, la selección de textos en Internet, esto es, enlaces a otros servidores que recogen recursos Web seleccionados de entre los pertenecientes a otras "bibliotecas digitales" (esta es la filosofía de la formación de colecciones digitales en *subject gateways*) y por otro, textos digitalizados pertenecientes al Servicio de publicaciones de la Obra Social Bancaja en formato PDF (que responde a la filosofía de creación de bibliotecas digitales creadoras y poseedoras de su colección digital).

³⁵ Cfr. *infr.* 8.2.



Fig. 59. Ciberoteca. Bancaja (pantalla principal)

- 3) Obras completas de Gonzalo de Berceo <<http://www.geocities.com/urunuela1/berceo/berceo1.htm>> En realidad se trata de un sitio especializado, un portal temático en sentido amplio, sobre dicho autor medieval, que incluye, entre otras cosas el acceso a sus obras completas en formato electrónico. Este sitio es parte de una página personal <<http://www.geocities.com/urunuela1/urunuela.htm>> realiza por Pedro Benito y dedicada a un pueblo de la Rioja. Desde nuestro punto de vista no podría considerarse una biblioteca digital a pesar de la decisión del portal de la UNESCO de incluirla como "biblioteca digital".
- 4) Obras de Miguel de Cervantes. Aunque en el momento de acceso a este recurso (24/10/2001) no funcionaba el enlace a esta biblioteca, pensamos, a tenor de la descripción que hace de ella el Portal de la UNESCO³⁶ (*esta página contiene enlaces a la obra completa de Miguel de Cervantes, distribuida por categorías. Las obras están disponibles también en listados por orden cronológico*) se trata de un caso semejante al anterior de la obra de

³⁶ UNESCO Libraries Portal <http://www.unesco.org/Webworld/portal_bib> Preservation and Access initiatives : Access : Digital libraries : Collections : Spanish

Berceo: un conjunto de enlaces, probablemente, en este caso, a otros servidores, sobre la obra de Cervantes en formato HTML. El único componente de organización, parece ser la existencia de índices cronológicos para el acceso a las obras del principal exponente de la literatura española.

- 5) Biblioteca electrónica cristiana <<http://www.multimedios.org>>. Esta denominada, en este caso "biblioteca electrónica" es un proyecto peruano que conforma una especie de vortal con enlaces a textos sobre la Iglesia y el cristianismo. Incluye un índice alfabético por autores y un índice temático que clasifica, en diecinueve categorías, los documentos que contiene este sitio. Además tiene un buscador por palabras clave que permite la búsqueda booleana básica (AND, OR), basado en la indización total de los textos. La selección de "enlaces" <<http://www.multimedios.org/ial/index.html>> que incluye es también un directorio temático sobre la Iglesia en América Latina, basado en el principio de recuperación de información del *browsing* o navegación. En sentido amplio podríamos decir que se trata de una biblioteca digital, en tanto que aprovecha las oportunidades del hipertexto para organizar, de alguna forma, la información que alberga.
- 6) Biblioteca Virtual Miguel de Cervantes Saavedra <<http://cervantesvirtual.com>>. Es un complejo y riquísimo servicio de información —al cual dedicaremos por completo el capítulo 9 de esta tesis— que sí puede considerarse una verdadera biblioteca digital, que incluye otros servicios que magnifican su potencial como portal cultural y tiene una filosofía implícita de servicio de biblioteca universal³⁷.
- 7) Finalmente, también se recoge en esa misma categoría establecida por el portal de bibliotecas de la UNESCO, un enlace a Bibliotecas Digitales en español <http://www.galeon.com/fernabriz_00/BibliotESP.htm> que no es en sí

³⁷ Cfr. *infr.* algunas definiciones en torno al término "biblioteca virtual". Vid. Capítulo 9.

mismo una biblioteca digital sino un recurso que recopila enlaces a "bibliotecas digitales". Como se dice expresamente, recoge *algunas direcciones de paginas que nos suministran full text de libros en español o de autores españoles*. Tanto el tipo de enlaces que acopia como esa afirmación expresa, nos da una pista sobre lo que se entiende genéricamente por biblioteca digital: un acceso electrónico a libros en texto completo, sin ninguna consideración a la selección, organización y formación de colecciones.

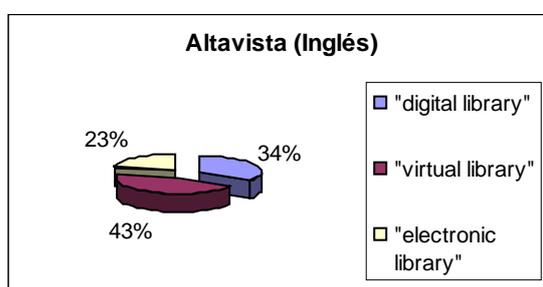
Como demuestra el ejemplo anterior, se utilizan³⁸ como sinónimos biblioteca virtual, biblioteca digital y biblioteca electrónica e incluso otras denominaciones parasintéticas e insólitas como "ciberoteca" o biblioteca VirtuArt. Sin embargo, en sentido estricto, hay matices reales que las diferencian o deberían diferenciarlas. La falta de univocidad designativa, complica aún más la determinación del concepto.

Si realizamos un muestreo en la propia WWW, utilizando la posibilidad de búsqueda por cadenas de caracteres que presentan los motores de recuperación de información en Internet, Altavista y Google, la realidad de la frecuencia de aparición es la siguiente:

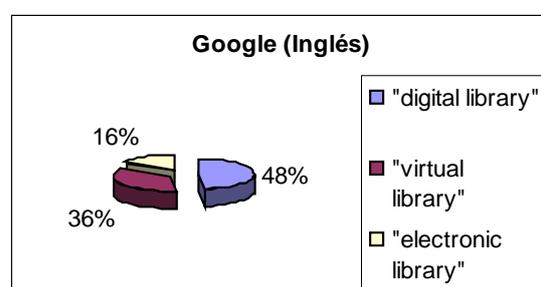
³⁸ De hecho, nosotros también intercambiamos la denominación por un lado, para enriquecer la expresión de este trabajo y por otro, porque se utilizan como sinónimos en la mayoría de los casos para denominar la realidad difusa de la promoción del acceso a los DLOs. Nos permitimos la parcialidad de hacerlo, a riesgo de contradecir las hipótesis que planteamos aquí, porque independientemente de que utilicemos "biblioteca digital, "biblioteca electrónica" o "biblioteca virtual", siempre nos estamos refiriendo a la realidad que defendemos en esta tesis de *conjunto organizado de DLOs*, que nos permite hablar de biblioteca, independientemente del calificativo escogido para patentar la realidad del acceso a través de la Red a esa información organizada y recuperable.

	Altavista	Google
"digital library"	295.105	670.000
"virtual library"	368.430	503.000
"electronic library"	193.815	228.000

Tabla 15. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en inglés (Altavista y Google)³⁹



Graf. 7. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en inglés (Altavista)



Graf. 8. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en inglés (Google)

Según los resultados de la frecuencia de aparición de los términos analizados en inglés, un 43% de las páginas que recoge Altavista sobre este tema, prefieren la denominación de "virtual library", mientras que en Google se recoge una mayor frecuencia (48%) del término "digital library". Otros estudios previos, análogos al muestreo que realizamos aquí, presentan datos semejantes. Por ejemplo, Bia y Pedreño⁴⁰, en un análisis realizado en julio de 2000 utilizando la búsqueda avanzada de Altavista (realizando las siguientes búsquedas "digital library OR digital libraries"; "virtual library" OR "virtual libraries"; "electronic library" OR "electronic libraries" e "internet library" OR "internet libraries") obtuvieron que el 56% de las

³⁹ Datos tomados de las búsquedas efectuadas el 06/11/2001.

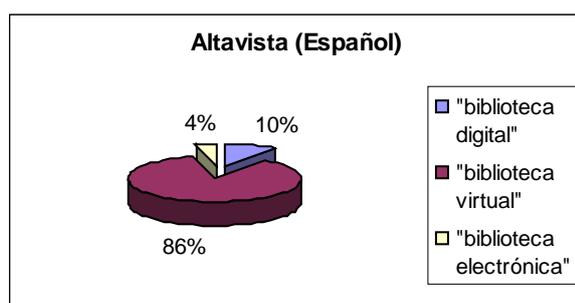
⁴⁰ Alejandro Bia, Andrés Pedreño. The Miguel de Cervantes Digital Library: The Hispanic Voice on the Web. *Literary and Linguistic Computing*, 2001, vol. 16, nº 1, p. 150, fig. 1.

páginas utilizaban la denominación de virtual library/ies, el 27% digital library/ies, el 14% electronic library/ies, y el 3% Internet library/ies. Los porcentajes son semejantes a los que hemos obtenido nosotros en Altavista, y los desvíos se deben fundamentalmente al momento del estudio y a que en nuestro caso sólo hemos contemplado la expresión absoluta en singular.

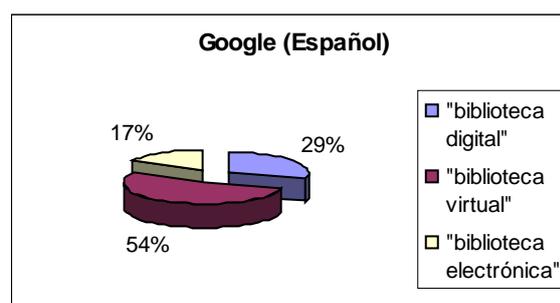
Si hacemos la misma búsqueda en español, obtenemos resultados semejantes con un claro predominio, en nuestra lengua, del término "biblioteca virtual" (un 86% en los resultados obtenidos en Altavista, y un 54% de las veces que aparece el término en las páginas que recoge Google en su base de datos):

	Altavista	Google
"biblioteca digital"	6.543	25.700
"biblioteca virtual"	56.895	47.800
"biblioteca electrónica"	2.583	14.900

Tabla 16. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en español (Altavista y Google)⁴¹



Graf. 9. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en español (Altavista)



Graf. 10. Impacto en la Web de los términos utilizados para referirse a la biblioteca digital en español (Google)

⁴¹ Datos tomados de las búsquedas efectuadas el 06/11/2001.

En cualquier caso, el análisis estadístico de la frecuencia de aparición de los términos en la propia Web sólo demuestra el reconocimiento de los tres conceptos. Ante esa reflexión, para sacar alguna conclusión más allá de lo elemental o evidente, sería preciso analizar todos y cada uno de los contextos en los que aparece cada una de las expresiones, así como todos los servicios que reciben una u otra denominación. Si la frecuencia de aparición o la "moda" designativa sirviesen como criterio para utilizar un término u otro, deberíamos denominar, a todas las tentativas informativas de carácter bibliotecario en la WWW, bibliotecas virtuales o digitales, sin embargo hay un porcentaje significativo —entre el 4 y el 17% en español, y en torno al 20% en inglés— que suscita la necesidad de clarificar también qué es una biblioteca electrónica.

Aunque en la práctica se tienden a intercambiar biblioteca digital, virtual y electrónica, como ya hemos apuntado, son varios los autores que diferencian con mayor o menor precisión estos términos incluyendo, en algunos casos, aspectos contradictorios. Resumimos a continuación algunas de las principales reflexiones en torno a la realidad que evoca cada una de estas tres denominaciones.

Una biblioteca electrónica es aquella que cuenta con sistemas de automatización que le permiten una ágil y correcta administración de los materiales que resguarda, principalmente en papel. Se concibe como la primera fase en la aplicación de las tecnologías de la información a los procesos documentales, esto es, lo que normalmente denominamos "automatización". López Guzmán⁴² añade a esta afirmación comúnmente aceptada, que la biblioteca electrónica *cuenta con sistemas de telecomunicaciones que le permitirán acceder a su información, en formato electrónico, de manera remota o local. Proporciona principalmente catálogos y listas de las colecciones que se encuentran físicamente dentro de un edificio.*

⁴²Clara López Guzmán. *Modelo para el desarrollo de bibliotecas digitales especializadas* [documento HTML]. México, D. F: UNAM, marzo de 2000. Disponible en: <http://www.bibliodgsca.unam.mx/tesis/tes7cllg/meta.html> (consultado el 1 de marzo de 2001).

Ontalba⁴³ sin embargo, señala un carácter más restrictivo en cuanto al acceso, diciendo que la biblioteca electrónica *da acceso local a información en formato electrónico, aunque no implica el uso de las telecomunicaciones ni elementos digitales*. La biblioteca electrónica que señala la primera autora es el modelo más convencional de todas las bibliotecas que han puesto sus OPACs en Internet (WebPACs), accesibles tanto dentro como fuera de la biblioteca, lo que implica una automatización del servicio pero no un acceso digital a los fondos. Sin embargo Ontalba constriñe el concepto de biblioteca electrónica a un nivel de acceso OPAC local.

A pesar de los contrastes que apuntamos en la aprehensión del concepto de biblioteca electrónica, la verdadera confusión surge a la hora de diferenciar biblioteca digital y virtual, e incluso de asimilar una realidad para ambas. Candy Schwartz⁴⁴ recoge sesenta y cinco definiciones de diferentes autores que, en algunos casos escogen para la designación el determinante "virtual" y en otros "digital", pero básicamente asimilan, con pequeños matices ambos conceptos. Una de las descripciones o definiciones genéricas más citadas en la literatura profesional⁴⁵, es la de la *Digital Library Federation* (DLF) que entiende la biblioteca digital (virtual) como una organización que proporciona los recursos, personal, etc. para seleccionar, estructurar, ofrecer acceso intelectual, interpretar, distribuir, preservar la integridad y asegurar la pervivencia, a través del tiempo, de colecciones de obras digitales, de tal

⁴³ José Antonio Ontalba y Ruipérez. Bibliotecas digitales españolas: análisis de las primeras realidades. En: *Jornadas Españolas de Documentación* (7. Bilbao. 2000). *Op. cit.*, p. 400-401.

⁴⁴ Candy Schwartz. *Definitions-LIS 530Z: Digital Libraries* [documento XHTML]. Boston: Simmons College, rev. 30 de marzo de 2000. Disponible en: <http://web.simmons.edu/~schwartz/530-defs.html> (consultado el 10 de febrero de 2001).

Todas las definiciones que recoge Schwartz, profesora del *Simmons College*, fueron recopiladas por sus alumnos entre toda la literatura profesional sobre el tema, en el segundo semestre de 1999. No sólo la cantidad, sino también la diversidad de conceptos asociados al término "biblioteca digital", es una muestra de la complejidad que señalamos.

⁴⁵ Sarah B. Watstein, Pascal V. Calarco, James S. Ghaphery. Digital library: keywords. *Reference Services Review*, 1999, vol. 27, nº 4, p. 349.

forma que puedan estar rápida y económicamente disponibles para ser usadas por una comunidad determinada.

En un intento más escrupuloso de diferenciación entre las tres denominaciones fundamentales que reflejamos aquí, podemos decir que el concepto de biblioteca digital implica el uso de las telecomunicaciones, es decir, el acceso a la información en *forma remota e independiente del lugar y número de conexiones*⁴⁶ y eso la diferencia fundamentalmente de la biblioteca electrónica; da acceso a una colección de información digital que normalmente *se encuentra disponible en una sola localización*. Mientras, la biblioteca virtual suele estar formada por un conjunto de conexiones a varias fuentes de información recogidas en Internet⁴⁷. Según esto, la única diferencia entre una biblioteca virtual y una digital es que en la primera los fondos están distribuidos, mientras que en la biblioteca digital, la colección es local. No obstante, a pesar de la lógica que podría tener esta afirmación sobre el concepto de colección virtual o colección distribuida⁴⁸, en la realidad, encontramos bibliotecas en la red que invalidan esta diferenciación, como la Biblioteca Virtual *Miguel de Cervantes* <<http://cervantesvirtual.com>> que custodia en una misma localización, en un mismo servidor, todos los DLOs que produce.

Por otra parte, si adoptamos una postura técnica y estricta, subrayando además el oxímoron de la expresión, una biblioteca virtual sería sólo *aquella que hace uso de la realidad virtual para mostrar una interfaz y emular un ambiente que sitúe al usuario dentro de una biblioteca tradicional*⁴⁹. La biblioteca virtual según esto trataría de emular una biblioteca convencional no sólo como la entendemos aquí,

⁴⁶ Pedro Hípola, Benjamín Vargas-Quesada y José A. Senso. Bibliotecas digitales: situación actual y problemas. *El profesional de la información*, abril 2000, vol. 9, nº 4, p. 5.

⁴⁷ *Ibid.*

⁴⁸ *Vid. infr.* 8.2.2.

⁴⁹ López Guzmán. *Op. cit.*, <http://www.bibliodgsca.unam.mx/tesis/tes7c1lg/meta.html>

como concepto, sino también como lugar o un espacio. Para conseguir remedar sobre la pantalla del ordenador ese espacio bibliotecario deben utilizarse tecnologías multimedia de simulación de la tridimensionalidad como el lenguaje VRML (*Virtual Reality Modeling Language*) o su sucesor X3D (*eXtensible 3D*)⁵⁰ que permite representar pasillos, paredes, estanterías etc. como si se tratase de una biblioteca real. Son pocas las experiencias de desarrollo de mundos virtuales adaptados a las bibliotecas, no obstante cabe destacar el proyecto de Jack Colbert, Librarea⁵¹.

Para aclarar todas estas definiciones o aproximaciones a uno y otro concepto, nos parece especialmente acertada la reflexión de Tomás Saorín⁵². Este autor parte, como nosotros, de la biblioteca como *institución documental* e incluye en su reflexión, además de las tres posibilidades que hemos barajado aquí (biblioteca electrónica, digital y virtual), las denominaciones "biblioteca universal" y "biblioteca global", que si bien no son aplicables a toda la WWW, como dijimos, sí son muy utilizadas para calificar determinados proyectos de biblioteca digital. Saorín diferencia dos esferas: por un lado la biblioteca como organización de servicios y por otro, la biblioteca como idea cultural. Según esta división, los adjetivos electrónico y digital pertenecen a la esfera de la biblioteca como organización, y los de universal y global a la abstracción de la biblioteca como ideal cultural. Universal y global son determinantes que conciben a la biblioteca como un ideal, situando la

⁵⁰ El lenguaje X3D fue presentado el 7 de agosto de 2001 como el sucesor de VRML para la representación tridimensional de la realidad en el ordenador. Se ha desarrollado en el seno del Consorcio Web 3D con el apoyo de algunas de las principales compañías de navegadores de "mundos virtuales" como blaxxun, Nexternet, OpenWorlds and ParallelGraphics. Están trabajando para convertir X3D en un estándar ISO así como en colaboración con el W3C y el MPEG-4 para integrar X3D con XML. Para más información sobre este nuevo estándar, Vid. <http://www.Web3d.org/x3d>

⁵¹ Para poder ver esta biblioteca virtual es necesario descargarse el conector de ActiveWorlds <<http://www.activeworlds.com/products/download.asp>> y acceder al "World" denominado Librarea. Más información sobre Librarea, Vid. <http://www.swrlss.org/aw1.html>

⁵² Tomás Saorín. *Los Portales Bibliotecarios: Gestión automatizada de bibliotecas y Sistema de Información Público en Internet* [documento PDF]. Murcia: Universidad, Grupo de Tecnologías de la información. Disponible en: <http://gti1.edu.um.es:8080/portales/FGSR-portales-Bibliotecas.PDF> (consultado el 3 de octubre de 2001).

conceptualización en un marco tecnológico difuso, en la que lo más importante es el alcance o ámbito del sistema de información. El uso de "digital" o "electrónico" para este autor, *se centra en la incidencia de cierta forma de la tecnología en la actividad de la institución biblioteca.* En esta dicotomía, la biblioteca virtual, *comparte elementos de las dos esferas, la organizacional-tecnológica y la abstracta-alcance.* Así una biblioteca virtual sería —desde la concepción filosófica de la biblioteca como idea cultural— aquella que trata de ser un servicio mundial, de dar un acceso universal y por ello, tiene en cuenta al usuario global, y a la vez se vale de la tecnología y de las telecomunicaciones para realizar la organización de sus servicios y colecciones⁵³.

Aunque hemos reflejado una parte de la dialéctica en torno a la denominación de las nuevas bibliotecas digitales surgidas en la Web, y de que hemos apuntado una diferenciación, si hiciésemos un análisis exhaustivo de todos los servicios y sitios en Internet que reciben una u otra denominación, en muy pocos casos la expresión formal que se utiliza para designarlos coincidiría con los matices que hemos señalado aquí. Así, utilizaremos las expresiones como sinónimas dando preeminencia, por su carácter inclusivo, a la de "biblioteca digital" en cuyo concepto confluyen, siguiendo a Borgman⁵⁴, una perspectiva de investigación desde la cual las bibliotecas digitales son contenidos recopilados y organizados en beneficio de los usuarios, y una perspectiva práctica donde son instituciones u organizaciones que proporcionan servicios de información de forma digital.

Para nosotros, desde la hipótesis que manejamos en esta tesis de la utilidad de los metadatos para optimizar la recuperación de información, una biblioteca digital no es más que, como hemos dicho: un conjunto *organizado* de DLOs, siempre y

⁵³ Esta es por ejemplo, la justificación que encontramos a la denominación de la Biblioteca Virtual *Miguel de Cervantes*, sobre la que centraremos nuestra investigación en el capítulo siguiente. *Vid.* 9.2.1., Fig. 62. Expresión gráfica del concepto de Biblioteca Virtual aplicado a la BVC.

⁵⁴ Christine L. Borgman. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, London: The MIT Press, 2000, p. 51.

cuando, tanto su contenido como su acceso, sean digitales. Los metadatos sirven para organizar los recursos de información electrónica, facilitando su acceso y recuperación, independientemente de que estén almacenadas en un mismo servidor o que estén distribuidos en distintos nodos en la Red. Como ya dijimos, todos los tipos de información, en formato digital (texto, datos, imágenes, vídeo, música, etc.) pueden estar presentes en estas bibliotecas, siempre que se den las condiciones necesarias para lograr su organización y difusión.

Como dice el profesor Moreiro⁵⁵ *las bibliotecas siempre se han preocupado de facilitar el acceso referencial y material a la información. Su objetivo no ha cambiado. A la hora de definir la nueva situación se han buscado denominaciones sinónimas como biblioteca virtual, biblioteca electrónica, biblioteca digital, biblioteca sin paredes...* y otras denominaciones espurias e insólitas, que tratan de adjetivar y calificar un viejo concepto al que se le ha incorporado o al menos, se le ha yuxtapuesto un nuevo medio, el medio electrónico y el soporte intangible de la WWW. La idea que hace prevalecer el valor de la biblioteca en sí misma, como determinamos en esta tesis "la biblioteca como concepto", la reafirman categóricamente, Braude y Wood que dicen que término biblioteca virtual o digital no es un oxímoron sino que es redundante, si no hemos importunado nuestras bibliotecas tradicionales denominándolas *bibliotecas de arcilla* o *bibliotecas de papiros enrollados, por qué ahora tenemos de llamarlas bibliotecas digitales*⁵⁶. Así pues, en todas esas denominaciones —como continúa Moreiro— *lejos de que los adjetivos se sitúen como modificadores de un núcleo, parecen haberse puesto en su*

⁵⁵ José A. Moreiro González. La distribución de los contenidos en la nueva sociedad informacional. En: *La Sociedad de la Información: política...* *Op. cit.*, p. 312.

⁵⁶ Robert Braude & Samuel Wood. Virtual or Actual: The Term Library is Enough. *Bulletin of the Medical Library Association*, 1999, nº 87, p. 87 [sic.]. Concepto recogido en: Candy Schwartz. *Definitions-LIS 530Z...* *Op. cit.*, <http://web.simmons.edu/~schwartz/530-defs.html> (consultado el 10 de febrero de 2001).

lugar, haciendo perder al término biblioteca la posición central, y dando al medio más importancia que al objeto⁵⁷.

8.2. Bibliotecas virtuales: del oxímoron del término a la realidad práctica

Como hemos reflejado, el propio término "biblioteca virtual" es un oxímoron. Greenberg⁵⁸ atestigua esta afirmación extendiéndola también a la denominación de "biblioteca digital", radicalizando incluso el discurso en torno al oxímoron, al afirmar que *si una biblioteca es una biblioteca, no es digital; si una biblioteca es digital, no es una biblioteca*⁵⁹. Sin embargo, en este apartado vamos a tratar de atestiguar como la implementación de bibliotecas digitales, perfectamente organizadas a través estándares para la recuperación de información (metadatos), convierten el oxímoron de la denominación en un concepto de realidad práctica y en un índice de contemporaneidad de la gestión bibliotecaria tradicional.

En la concepción de la que partimos en esta investigación para aplicar metadatos a bibliotecas digitales que suponen contextos finitos, abarcables y locales, frente a lo infinito, indefinido y global de los sistemas de recuperación destinados a la recuperación genérica en toda la Red, resulta especialmente significativa la caracterización de López Guzmán⁶⁰. Esta autora define la biblioteca digital como *un repositorio de acervos y contenidos digitalizados, almacenados en diferentes*

⁵⁷ José A. Moreiro González. *Loc. cit.*

⁵⁸ D. Greenberg. Camel Drivers and Gatecrashers: Quality Control in the Digital Research Library. En: *The Miriage of Continuity: Reconfiguring Academic Information Resources for the 21st Century*. B. L. Hawkins and P. Battin, eds. Washington, DC: Council on Library and Information Resources and Association of American Universities, 1998, p. 106. Citado en: Christine L. Borgman. *Op. cit.*, p. 38.

⁵⁹ Cfr. Robert Braude & Samuel Wood. Virtual or Actual: The Term Library is Enough. *Bulletin of the Medical Library Association*, 1999, n° 87, p. 87 [sic.]. Citado en: Candy Schwartz. *Definitions-LIS 530Z...* *Op. cit.*, <http://web.simmons.edu/~schwartz/530-defs.html>

⁶⁰ López Guzmán. *Op. cit.*, <http://www.bibliodgsca.unam.mx/tesis/tes7cllg/meta.html>

formatos electrónicos... generalmente, son bibliotecas pequeñas y especializadas, con colecciones limitadas a sólo algunos temas. Los metadatos, tal y como se conciben hoy, como hemos reflejado en las páginas precedentes de esta investigación, servirán para organizar "pequeñas"⁶¹ bibliotecas o servicios de información digital finitos, con límites establecidos y preferiblemente de una materia concreta. El hecho de estas bibliotecas sean temáticas responde a la tendencia general a la verticalización de la información en la Red, que señalábamos en el capítulo anterior. Además la especialización temática de un servicio de información o biblioteca virtual, permitirá adoptar planteamientos de recuperación de información fundados no sólo en metadatos basados en el recurso, sino también metadatos basados en materias (como la utilización de RDF para construir ontologías que establecen un vocabulario común de la terminología o lenguajes controlados como tesauros, que servirán para una mayor consistencia de la recuperación por materias a través de metadatos descriptivos del contenido⁶²).

De todas las definiciones que recoge Candy Schwartz en el recurso que ya hemos citado⁶³ deduce los objetivos que idealmente deben tener las bibliotecas digitales, que delimitan el concepto, a la par que reafirman nuestra teoría de aplicación de metadatos a sus colecciones de DLOs (bien sean éstos producto de la digitalización *ad hoc*, o bien producidos originalmente en formato digital y seleccionados a posteriori):

⁶¹ Nótese las comillas en el adjetivo "pequeñas", ya que el tamaño de estas bibliotecas digitales, aunque frente al tamaño de toda la Red permite tildarlas de "pequeñas", sus colecciones, en número de DLOs que contienen o a los que dan acceso, son mucho mayores que los fondos, en número de documentos (o registros), en cualquier biblioteca tradicional.

⁶² Como expresamos en el capítulo 2 (*Cfr.* 2.1), si bien no están excluidos *ex professo* de esta tesis, los metadatos basados u orientados a las materias, sobre todo, su aplicación práctica concreta, trasciende los objetivos inmediatos de este trabajo, pero genera nuevas perspectivas de investigación en metadatos en la misma línea de aplicabilidad a bibliotecas "verticales" temáticamente.

⁶³ Candy Schwartz. *Definitions-LIS 530Z... Op. cit.*, <http://web.simmons.edu/~schwartz/530-defs.html>

- Servir a una comunidad definida o conjunto de comunidades.
- No debe ser una entidad aislada.
- Deben estar apoyadas por una estructura organizacional lógica y organizada.
- Incorporar tanto el aprendizaje como el acceso.
- Aprovechar al máximo tanto los recursos humanos (bibliotecarios) como tecnológicos.
- Proporcionar un acceso rápido y eficaz contemplando múltiples modos de acceso.
- Facilitar el acceso libre (quizás sólo a una comunidad específica).
- Poseer y controlar sus recursos.
- Formar amplias colecciones, que sean persistentes en el tiempo y que estén bien organizadas y gestionadas. Dichas colecciones deben contener además, diferentes formatos, objetos y no sólo sus representaciones⁶⁴, objetos que pueden no conseguirse de otra forma, y algunos objetos (DLOs) que sean digitales *ab origine*.

Esta característica, la formación de colecciones persistentes y organizadas, es el factor fundamental que nos lleva a defender la idea de biblioteca digital basada en metadatos, donde la búsqueda y recuperación de información, funciones últimas de un servicio de información, serán más coherentes que la búsqueda sobre el texto electrónico completo. Obviamente, esa organización de la colección podría basarse también en cualquier otro estándar de carácter bibliográfico, sin embargo, nuestra teoría de los metadatos en sentido estricto (como normas para la estructuración de la

⁶⁴ Aunque Schwartz utiliza aquí el término inglés *representations*, se refiere al concepto de información referencial, es decir, a la referencia bibliográfica, o mejor dicho en el ámbito de los metadatos el *subrogado* del DLO. Cfr. Glosario, *surrogate*, *subrogate*.

información electrónica y como una nueva solución de carácter bibliotecario e informático para un nuevo problema de gestión de información electrónica), unida a la tendencia general de la Web Semántica futura, nos hacen pensar que los metadatos son la mejor forma de organizar una biblioteca digital.

A continuación vamos a tratar las dos realidades de biblioteca digital que reconocemos en esta tesis:

- Por un lado, aquellas que, además de dar acceso a una colección digital a través de un sistema de búsqueda, poseen los documentos "digitalizados", que son almacenados en un sistema, normalmente local o centralizado. A este primer tipo de bibliotecas las hemos denominado poseedoras, y también productoras en muchos casos, de la colección digital a la que dan acceso. En ellas los metadatos deben plantearse como una estructuración pre-publicación de la información de los distintos DLOs, creados en procesos más o menos complejos de digitalización⁶⁵.
- Por otro lado, las bibliotecas digitales que fundan su servicio de búsqueda y recuperación de información en la constitución de un catálogo que refleja una colección virtual distribuida. Los metadatos se crearán pues según el modelo de representación post-publicación. Denominamos este tipo de bibliotecas digitales, siguiendo la terminología más usada internacionalmente, *subject gateways*, pasarelas temáticas o bibliotecas virtuales temáticas⁶⁶.

Otra realidad comúnmente reconocida en el ámbito de las bibliotecas digitales son las bibliotecas híbridas como modelo de transición entre la biblioteca tradicional

⁶⁵ Esta es la visión que defenderemos en el capítulo siguiente, al ser la Biblioteca Virtual Miguel de Cervantes un paradigma de las características que, como expondremos a continuación (8.2.1) concurren en este tipo de bibliotecas digitales. *Cfr.* Capítulo 9.

⁶⁶ *Cfr.* Glosario, *subject gateway/information gateway*.

y la biblioteca enteramente digital, que denota un estado de interrelación entre la tecnología y la visión institucional tradicional de la biblioteca, pero el concepto en sí mismo está tan difuso como el de biblioteca digital/virtual/electrónica. El término de "biblioteca híbrida" surge en el seno del proyecto eLib y recoge la visión de los proyectos de las bibliotecas digitales académicas del Reino Unido, sin embargo, las técnicas de hibridación de las colecciones son anteriores a la propuesta del término y van, como hemos apuntado desde la catalogación convencional de recursos electrónicos remotos en formato MARC a través del campo 856, integrando documentos y DLOs en el catálogo, la digitalización de colecciones impresas, a la inclusión de una página de enlaces seleccionados dentro de las páginas Web que responden a instituciones bibliotecarias y que suponen la asunción del valor de la información electrónica. Sin embargo, esta situación de transición, esta realidad bibliotecaria, la trataremos en el apartado siguiente al hablar de la interoperabilidad necesaria para dar acceso a un servicio de información bibliotecaria global interrelacionado⁶⁷.

No obstante, muchos autores conciben la biblioteca híbrida *como un esfuerzo que trata convertir paulatinamente los recursos de las bibliotecas digitales en servicios de producción y preservación digital*⁶⁸. Así el director del programa eLib⁶⁹ dice que estas bibliotecas híbridas deben integrar el acceso a cuatro tipos de recursos (tradicionales, recursos digitalizados, nuevos recursos creados en formato digital *ex novo*, y recursos digitales normalizados) de tal forma que se armonicen todo tipo de tecnologías en el contexto de una biblioteca, al mismo tiempo que servirá para experimentar la integración de sistemas y servicios en contextos electrónicos e impresos. La biblioteca híbrida es un servicio de información que refleja la evolución

⁶⁷ Cfr. *infr.* 8.3.2.

⁶⁸ V. gr. Chris Rusbridge. Towards the Hybrid Library [documento HTML]. *D-Lib Magazine*, July/August 1998, vol. 4, nº 7/8. Disponible en: <http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html> (consultado el 21 de septiembre de 1999).

⁶⁹ *Ibid.*

de la biblioteca tradicional (materiales tradicionales) a la biblioteca digital *del futuro* (tal y como la concebimos aquí, formada por DLOs representados por metainformación).

Ahora nos referiremos sin embargo a las bibliotecas que podemos considerar digitales en sentido estricto, es decir aquellas que están constituidas por DLOs normalizados. Para ello nos basamos en las dos proposiciones fundamentales de la recuperación de información en bibliotecas con las que cerrábamos el capítulo anterior (Fig. 58), que determinaba que el proceso de recuperación de información implica crear un límite definido y que los usuarios estén familiarizados con el propósito de la información buscada. Estas dos proposiciones nos hacen delimitar la operatividad de los metadatos en bibliotecas formadas por conjuntos abarcables de información que reflejamos a continuación.

8.2.1. Bibliotecas productoras/poseedoras de su colección digital

Bajo la denominación de bibliotecas virtuales/digitales poseedoras y productoras de su colección digital enmarcamos aquellas que se basan en la formación de sus colecciones en la digitalización y almacenamiento en formato electrónico de materiales convencionales, es decir, en trasuntos electrónicos o representaciones digitales de tales materiales. La motivación para la creación de estas bibliotecas puede ser desde la preservación de documentos originales a través de la facilitación del acceso a sus contenidos en formato digital, a la posibilidad de multiplicar el acceso simultáneo, pasando por la digitalización y organización de materiales descatalogados, la difusión de colecciones únicas o simplemente el deseo de proporcionar un acceso global al patrimonio cultural⁷⁰.

Su producción y almacenamiento es local, pero tiene una proyección de acceso global. La colección es real, una realidad más que digital, digitalizada, de la que

⁷⁰ Cfr. *infr.* 8.3.

podemos destacar las siguientes características, distinguiéndolas, por un lado, del paradigma de la biblioteca tradicional y por otro, de las bibliotecas digitales distribuidas:

- a) En vez de almacenar libros, almacena documentos entendidos como objetos de información digital (DLOs): toda una suerte de textos, vídeos, imágenes, etc. en distintos formatos de codificación numérica.
- b) Nos referimos a ellas también como "contextos finitos/abarcables" de información digital porque implican la digitalización (y creación de un sistema digital de información con ellos) de un número determinado de documentos, de una época, de un tipo de información o de una materia que permitirá diseñar un esquema de metadatos para la recuperación de información del material/es definidos que recopile.
- c) Los DLOs se almacenan en un servidor/es dentro de un marco institucional normalmente de forma centralizada, pero también puede ser un almacenamiento cooperativo de distintas instituciones implicadas en un determinado proyecto de digitalización.
- d) El almacenamiento de los objetos de información que componen la colección se hace de tal forma que permita la manipulación, la presentación y el acceso, que implique una forma de visualización y manipulación distinta a la versión convencional del documento. Estas formas de presentación y acceso distintas se consiguen: por un lado, a través de los lenguajes de marcado aplicados a la información textual y por otro, a través de la creación y gestión de metadatos para la recuperación.

En el sentido que las analizamos aquí deben incorporar paradigmas de metadatos bibliográficos para la representación interna de los distintos objetos de información, así como para fundar la recuperación externa.

- e) La creación de este tipo de bibliotecas digitales implica procesos de planificación exhaustivos tanto desde el punto de vista informático (hardware, software, almacenamiento, protocolos, etc.), como bibliotecario (determinación del modelo de metadatos, nivel de detalle de la descripción, etc.), así como otras cuestiones relacionadas con los derechos de autor o la propiedad intelectual de los recursos digitalizados⁷¹. Sin embargo entendemos que la formación de "catálogos" así como su acceso y opciones de búsqueda es el aspecto más visible de estas bibliotecas y por ello, el éxito de las mismas dependerá de la correcta extracción y/o asignación de metadatos.
- f) Además del cambio de paradigma en la organización y recuperación de la información que advertimos —y demandamos—, la creación de estas bibliotecas productoras de su colección, supone también un cambio en los flujos de trabajo bibliotecarios y una nueva concepción de la biblioteca (digital) como editora-creadora y gestora de los documentos digitalizados que recoge y como un nuevo sistema socio-técnico de información.

El hecho de que la biblioteca adopte al mismo tiempo las funciones de edición y gestión de su colección digital potencia, en nuestra opinión, la concepción tradicional de la biblioteca como sistema de Berta Enciso⁷², donde la biblioteca (tradicional) es la base del proceso educativo e investigador. Para Berta Enciso la biblioteca convencional entendida como sistema, está formada por dos subsistemas a

⁷¹ Todo el proceso de planificación e implementación de una biblioteca digital de estas características, cuyo detalle sobrepasa los objetivos de esta investigación está perfectamente descrito en: [Peter Noerr]. *The Digital Library Toolkit*. 2nd ed. Palo Alto: Sun Microsystems, March 2000, p. 45-110.

Para una visión más breve y centrada al desarrollo de un sistema de metadatos como núcleo de la biblioteca digital, *Vid.* Jean Hudgins, Grace Agnew, Elizabeth Brown. *Op. cit.*, p. 42-53, y un ejemplo de planificación e implantación aplicado a la *Georgia Tech Library and Information Center*: p. 54-59.

⁷² Berta Enciso Carvajal. *La biblioteca: bibliosistemática e información*. 2^a ed. México: Colegio de México, 1997 [documento HTML]. Alicante: Biblioteca Virtual Cervantes, 1999-2001. Disponible en: <http://www.cervantesvirtual.com/servlet/SirveObras/079596188953519752324333> (consultado el 20 de mayo de 2001).

través de los cuales lleva a cabo sus funciones y tareas: a) el subsistema informativo y administrativo, que se encarga de la planificación y evaluación de las bibliotecas; y b) el bibliográfico, que comprende tanto los servicios técnicos (adquisición, proceso, selección y eliminación de fondos) como los servicios públicos (préstamo, difusión, recuperación, mantenimiento y consulta de la colección). En las bibliotecas digitales se produce una reformulación de estos subsistemas.

- Por un lado, se crea un nuevo subsistema que podemos denominar "subsistema editorial" que comprenderá todos los procesos de creación de la unidad documental de la biblioteca (documento entendido como objeto de información), y que implica todas las actividades de digitalización, marcado estructural y publicación del DLO en la Web⁷³.
- Por otra parte, el subsistema informativo de Enciso, conllevará ahora, la planificación de la biblioteca desde el punto de vista de los flujos de trabajo de producción y se le añade además a la planificación, una dimensión informática fundamental.
- Finalmente, en el subsistema bibliográfico, desaparecen actividades y conceptos como el de adquisición, eliminación de fondos o préstamo que dejan de tener sentido en el nuevo entorno digital. Sin embargo se reactiva la importancia de la recuperación y se multiplica la consulta. El núcleo de este sistema bibliográfico, según el planeamiento que defendemos en esta tesis, será la asignación de metadatos a los DLO producidos, que, en este caso, estará íntimamente relacionado con el subsistema editorial.

⁷³ Es preciso recordar aquí que el DLO, el documento entendido como objeto de información, además de ser el fundamento de los metadatos puede también considerarse la unidad "bibliográfica" de una biblioteca digital. El fondo de una biblioteca digital está constituido por un conjunto de textos, imágenes, información sonora, vídeo, etc. convenientemente digitalizados y adaptados para la WWW. El subsistema editorial que reconocemos aquí se encarga de una función nueva pero básica en esta nueva concepción de biblioteca: crear el DLO.

Esta visión de la biblioteca digital forma un entramado más complejo y difuso que el que reconocía Berta Enciso para la biblioteca tradicional, donde se entremezclan aspectos sociales, económicos y culturales asociados al desarrollo de estas bibliotecas digitales que la convierten en un nuevo entramado socio-técnico y en un sistema fruto de los cambios estructurales de la Sociedad de la Información, que resumimos en la representación gráfica a continuación (Fig. 60).

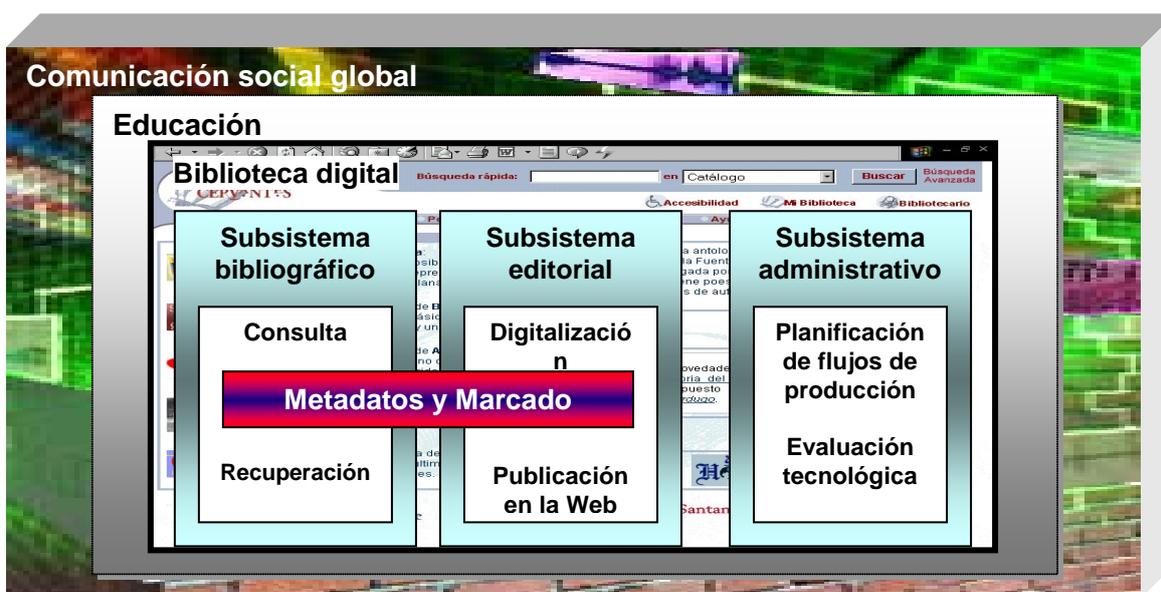


Fig. 60. Reformulación de la biblioteca como sistema en el mundo digital (Bibliotecas productoras-poseedoras de su colección digital)

La creación, almacenamiento y gestión de metainformación para la recuperación se debe realizar en estas bibliotecas, junto con el proceso de marcado, constituyendo parte de él, de tal forma que se puede cumplir el principio general de que los metadatos deben de asignarse por el propio creador del documento digital para la Web⁷⁴. Entendemos por creador del documento no el que convierte el documento en átomos a bits, sino el que le da a éste una estructura semántica a través del marcado, donde los metadatos deben ser, parte asociada, independientemente del modo de almacenamiento de éstos, en la que se fundará la recuperación.

⁷⁴ Cfr. 4.1. y 9.2.2.

La ARL (*Association of Research Libraries*) recoge en su base de datos de iniciativas digitales (DID⁷⁵), cuatrocientos diecinueve proyectos de digitalización que podrían entenderse, con más o menos rigor, dentro de este tipo de bibliotecas, pero los orígenes de estas bibliotecas productoras de sus colecciones digitales se fundan en la *Digital Library Initiative* (DLI) del gobierno de los Estados Unidos, que incluyó la investigación en bibliotecas digitales en el programa HPCC (*High Performance Computing and Communications*⁷⁶). Esta iniciativa, incoada en 1994 bajo los auspicios de la NSF, la DARPA y la NASA fue la respuesta a la demanda de sistemas y servicios de información de calidad, capaces de proporcionar un acceso coherente a grandes bases de datos a distintas comunidades de usuarios. La DLI es un proyecto de investigación para incrementar las formas de producción, almacenamiento y organización de información digital de tal forma que facilite la búsqueda, recolección y procesamiento de tales materiales electrónicos a través de redes de comunicaciones. La DLI se ha desarrollado en dos periodos (DLI1: 1994-1998 y DLI2: 1999-2004) y financia proyectos multidisciplinarios de bibliotecas digitales. En la primera fase (1994-98), los proyectos estaban liderados por seis universidades que iniciaron sus investigaciones particulares en bibliotecas digitales dando lugar a múltiples servicios, subproyectos y avances tecnológicos aplicables. La mayoría de ellos se basan fundamentalmente en la digitalización y recuperación distribuida⁷⁷ de diferentes tipos de materiales (las colecciones pueden estar formadas

⁷⁵ ARL Digital Initiatives Database <<http://db.arl.org/did/review.html>>. Esta base de datos no se actualiza desde el 30 de octubre de 1998.

⁷⁶ El programa de Informática y Comunicaciones de Alto Rendimiento (HPCC), es uno de los baluartes de la Infraestructura Nacional de Información (NII) americana. Vid. *HPCC FY 1995 Implementation Plan* [documento HTML]. National Coordination Office for Computing, Information and Communications, 26 de mayo de 1994. Disponible en: <http://www.ccic.gov/pubs/imp95> (consultado el 10 de mayo de 1999).

⁷⁷ El carácter distribuido de la recuperación en estos proyectos las aproxima a las bibliotecas virtuales temáticas, cuya colección está repartida en distintos servidores, que trataremos en el apartado siguiente (8.2.2.), sin embargo los incluimos dentro de las bibliotecas productoras de su información digital porque, en la mayoría de los casos, la fase inicial del proyecto incluye la digitalización y almacenamiento (aunque éste sea distribuido) de documentos tradicionales (texto, vídeo, imagen, etc.).

por texto, imágenes, mapas, audio, vídeo, o combinaciones multimedia de todos ellos):

- La Universidad de California en Berkeley cuyo objetivo era la creación de una la *Environmental Electronic Library* <<http://elib.cs.berkeley.edu>> que contemplara el ciclo de vida de la información académica en aras a facilitar un modelo descentralizado de autopublicación y recuperación de ese tipo de información. Basa su recuperación de información en un catálogo bibliográfico de segunda generación denominado Cheshire II.. Este proyecto tiene relación con CERES que mencionamos al hablar de los metadatos basados en materias⁷⁸.
- La Universidad de California en Santa Bárbara <<http://www.alexandria.ucsb.edu>> que lidera el proyecto, del que ya hemos hablado en otras partes de esta investigación, *ADL (Alexandria Digital Library)*, un sistema de información geográfica espacialmente referenciada, descrita y recuperada.
- La Universidad de Carnegie Mellon que lidera el proyecto Informedia <<http://www.informedia.cs.cmu.edu>> que incluye toda una investigación para la creación de bibliotecas digitales de vídeo digital y multimedia.
- La Universidad de Illinois en Urbana-Campaign, cuyo proyecto tratada de crear una infraestructura de biblioteca digital para hacer búsquedas eficaces de contenidos técnicos en Internet entre artículos de revistas de física, ingeniería e informática realizadas en SGML que están disponibles en la Red antes imprimirse, que denominaron *Building the Interspace* <<http://dli.grainger.uiuc.edu/idli/idli.htm>>.
- La Universidad de Michigan cuya investigación en bibliotecas digitales <<http://www.si.umich.edu/UMDL>> se basa en la creación de agentes de software para la búsqueda (agentes de colección, asociados a cada colección determinada,

⁷⁸ Cfr. 6.4. *California Environmental Resources Evaluation System*: <http://ceres.ca.gov>

- agentes intermediarios que permiten la búsqueda entre colecciones distribuidas y agentes de interfaz de usuario para analizar las necesidades de información de los usuarios). Su investigación en el ámbito de los metadatos se refiere a metadatos basados en materia a través de un catálogo basado en un modelo de ontología formal de relaciones bibliográficas⁷⁹. Dentro de los múltiples proyectos relacionados⁸⁰ con la UMDL, nos parece interesante destacar, en la categoría de bibliotecas productoras-gestoras de su colección, JSTOR (*Journal Storage Project*) que se encarga de digitalizar y organizar los principales títulos de revistas sobre ciencias sociales y humanidades. En este caso, aunque la biblioteca es editora de los DLOs, la producción se hace de forma descentralizada en tres centros (Nueva York, Anne Arbor y Princeton). Proporciona las imágenes de los artículos a los usuarios, sin embargo, las búsquedas se realizan sobre el texto completo digitalizado. Con una forma semejante de producción (imágenes + texto buscable marcado en XML) se realiza el proyecto MOA (*Making of America*) en colaboración con la Universidad de Cornell que digitaliza libros y artículos de la historia americana del siglo XIX.
- La Universidad de Standford, cuyo proyecto de bibliotecas digitales <<http://www-diglib.stanford.edu>> se centra en la investigación sobre el acceso integrado a servicios de información heterogénea en Internet desarrollando un protocolo específico denominado InfoBus y basado en CORBA, así como la interacción de este protocolo con otros que, como Z39.50, permiten el acceso a otras bases de datos bibliográficas.

⁷⁹ Esta estructura de metainformación dependiente del contenido se describe y analiza en: Peter C. Weinstein. *Ontology-Based Metadata: Transforming the MARC Legacy*. En: *ACM Conference on Digital Libraries (3.1998. Pittsburg)*. Ian Witten, Robs Akscyn, Frank M. Shipman, III, eds. New York: ACM, 1998, p. 254-263. *Vid. Figs.*, p. 258.

⁸⁰ Para un listado completo de estos proyectos, *Vid.* <http://www.si.umich.edu/UMDL/otherdiglib.html>

La II fase de la iniciativa americana de bibliotecas digitales (1999-2004) revela el éxito de la primera, e implica el reconocimiento de la importancia de estos servicios de información (bibliotecas digitales) en la Sociedad de la Información, al haberse concedido una dotación presupuestaria exorbitante tanto para la continuidad de los proyectos iniciados en la primera fase como para la incorporación de otros nuevos. Además del mecenazgo de la NSF, DARPA y la NASA, se han incorporado otras importantes instituciones como la Library of Congress, la NLM, o la Smithsonian Institution, entre otras. Destaca la inclusión en la DLI de doce proyectos internacionales protagonizados por Universidades públicas americanas y otras instituciones, principalmente británicas⁸¹.

Desde el punto de vista temático, las colecciones que desarrollan estos proyectos de la DLI², van desde la investigación en aplicaciones para la gestión de la información en bibliotecas digitales (como el proyecto de la Universidad de Arizona denominado, *High Performance Digital Library Classification Systems: From Information Retrieval to Knowledge Management*) a la creación de una biblioteca de textos cuneiformes (CDLI) liderado por la Universidad de California en Los Ángeles <<http://cdli.ucla.edu>> cuyos metadatos se crean en una base de datos diseñada de FileMaker Pro⁸² o la creación de una biblioteca digital multimedia de literatura judeo-sefardí, liderado por la Universidad de California en Davis <<http://flsj.ucdavis.edu/home>>. La diversidad de materias que refleja la generalidad de los proyectos, así como la uniformidad temática de cada uno de ellos evidencia, por un lado, el reflejo de uno de los principios de las políticas de información de ámbito internacional: facilitar el acceso al patrimonio cultural y a la información a través de Internet, y por otro, que estas bibliotecas digitales responden a contextos abarcables

⁸¹ Cfr. 7.2. Internet Scout Project. Vid. *International Meshes (IMesh): International Collaboration on Internet Subject Gateways*:<http://www.imesh.org>

⁸² Aunque en sentido estricto, los campos de una base de datos "bibliográfica" son metadatos, no se utiliza ningún modelo de estructuración de la metainformación a través del marcado.

de información digitalizada, a diferencia de las que analizaremos en el apartado siguiente, basadas en información originalmente digital.

Por otra parte, la casuística técnica que revelan todos estos proyectos indica una vez más, que las bibliotecas digitales son un campo de investigación creciente, donde los metadatos deben ser el núcleo de su sistema bibliográfico y por tanto, de la recuperación de información; y también pone de manifiesto que estas bibliotecas alimentan otras investigaciones de índole diversa, tanto desde el punto de vista de la aplicación de estándares como en la formación de bases de datos o de la creación de protocolos y *search engines* que faciliten la búsqueda entre colecciones heterogéneas.

Aunque no todas las bibliotecas digitales surgidas de las iniciativas vinculadas a la DLI, son productoras y poseedoras de sus colecciones y éstas, en muchos casos se encuentran distribuidas (bien la fase de producción, como en JSTOR, bien en el almacenamiento, o en ambas) todas ellas tienen un subsistema editorial o un sistema de conversión de documentos tradicionales en DLOs. Aunque las opciones de recuperación de información se basan en distintos modelos como en SGBD convencionales como FileMaker o Access, o en desarrollo de protocolos específicos como InfoBus de Standford, en todas ellas confluye la posibilidad, al menos *a priori*, de asignar los metadatos en el momento de creación del DLO, esto es, se puede diseñar un sistema de creación de metadatos asociado al momento inicial de marcado y/o publicación del documento en formato electrónico.

En España el paradigma de esta forma de biblioteca digital, por no decir el único modelo en el que concurren todos los subsistemas que reconocemos en dicha tipología —subsistema administrativo, bibliográfico y editorial—, es la Biblioteca Virtual *Miguel de Cervantes* a la que dedicaremos por completo el capítulo siguiente.

8.2.2. *Subject gateways* y el concepto de colección distribuida

Normalmente, el criterio físico de organización de una colección en la biblioteca tradicional es el soporte (libros, revistas, vídeos, etc.), y el criterio de organización del contenido intelectual, es la materia o tema que trata la publicación. Algo semejante ocurre en algunos servicios de Internet que conforman sistemas de recuperación de información basados en materias, basados en la formación y organización de colecciones (virtuales) distribuidas. Así, a parte de las bibliotecas que son, además de gestoras, productoras o editoras digitales de su colección, existe otro planteamiento de sistema de organización y recuperación de información en Internet que podemos considerar biblioteca digital/virtual, basándonos en los criterios que hemos establecido (selección y organización), y en las dos proposiciones básicas de la RI, (límite definido y propósito de la información buscada). Nos referimos los servicios de recuperación de información en Internet basados en materias *subject-based resource discovery services*, también denominados servicios de materias de calidad controlada (*quality controlled subject services*⁸³), e incluso, *clearinghouses*⁸⁴, o más común o simplemente *subject gateways* o *information gateways*⁸⁵.

⁸³ Así aparecen denominados en la clasificación de los sistemas de búsqueda que realiza Koch. Traugott Koch. *Browsing and Searching Internet Resources* [documento HTML]. Lund University Electronic Library, junio de 1994, rev. 11 de febrero de 1998. Disponible en: http://www.ub2.lu.se/nav_menu.html (consultado el 28 de julio de 1998). Según la clasificación de Koch, las *subject gateways* no serían más que un sistema especial dentro de los servicios de recuperación de información de Internet basados en la navegación por materias. Sin embargo, desde nuestro punto de vista, en este tipo de servicios concurren una serie de características que nos permiten considerarlos bibliotecas digitales, virtuales o quizás, mejor electrónicas, en el propio sentido que se les da en el marco del proyecto inglés eLib (*Electronic Libraries*).

⁸⁴ Diann Rusch-Feja. Subject-Oriented Collection of Information Resources from the Internet: A Clearinghouse Concept to Support Scientist in a German Research Institute. *Libri*, 1997, vol. 47, p. 1-24. Cfr. Glosario, *clearinghouse/s*

⁸⁵ Si bien el término más amplio de *information gateway* podría cubrir SRII de propósito general o global y carácter enciclopédico como Infomine o BUBL. Cfr. DESIRE. *Information Gateways... Op. cit.* <http://www.desire.org/handbook/print4.html> Vid. Glosario, *subject gateway/information gateway*.

El modelo informativo de estas bibliotecas digitales está muy próximo a la concepción de portales bibliográficos que señalábamos anteriormente (8.1.) por su vocación de incluir servicios de valor añadido a la información bibliográfica en texto completo, e incluso al concepto de biblioteca híbrida, que también hemos mencionado y que relacionaremos en el apartado siguiente (8.3.2.). Pero sobre todo se basan en el concepto creciente de colección distribuida o colección virtual. En el contexto distribuido de estas bibliotecas digitales, las colecciones se pueden definir, siguiendo a Lagoze y Fielding⁸⁶, desde una doble perspectiva: lógica y operacional. Desde la perspectiva lógica una colección es un *conjunto de criterios para la selección de recursos en un espacio de información amplio*. Desde la perspectiva operacional, una colección proporciona herramientas para la recuperación de información, y los recursos en esa colección son aquellos que se pueden encontrar directamente utilizando dichas herramientas de recuperación.

Esta amplitud del espacio de selección de recursos en la nueva concepción de las colecciones trasciende también a las bibliotecas tradicionales, y tiene relación con la idea difundida por las bibliotecas de la Universidad de Rochester⁸⁷, acerca de que los catálogos bibliotecarios han ampliado su papel tradicional y son un inventario de los recursos a los que la biblioteca tiene acceso, más que de lo que la biblioteca posee. Esta misma idea en torno al acceso más que a la posesión de los documentos, que subyace a la *nueva biblioteca* y a los nuevos catálogos, la refrenda el profesor Gómez⁸⁸ al decir que *las bibliotecas, aunque sigan dando acceso a los documentos*

⁸⁶ Carl Lagoze, David Fielding. Defining Collections in Distributed Digital Libraries [documento HTML]. *D-Lib Magazine*, November 1998, vol. 4, n° 11. Disponible en: <http://www.dlib.org/dlib/november98/lagoze/11lagoze.html> (consultado el 4 de febrero de 2001).

⁸⁷ *Cataloging Guidelines for Internet Resources*. Drafted by Electronic Publications Implementations Team [documento HTML]. New York: University of Rochester Libraries, septiembre 1997, rev. 18 de septiembre de 1998. Disponible en: <http://www.lib.rochester.edu/cat/catguide.htm> (consultado el 28 de junio de 2000). *Cfr.* Capítulo 4, nota 17.

⁸⁸ José Antonio Gómez Hernández. Legitimación y funciones de la biblioteca en el contexto de la sociedad digital. *Scire*, 1998, vol. 4, n° 2, p. 72.

impresos [...] deberán dar acceso a materiales que no sean de su propiedad, y sobre todo cada vez deberán proporcionar más recursos electrónicos, realizando sobre ellos tareas técnicas parecidas a las tradicionales. Esas tareas técnicas "parecidas" a las tradicionales son obviamente los metadatos, y fruto de esa necesidad ineluctable de proporcionar acceso a los recursos electrónicos, es lo ha llevado a la creación de bibliotecas virtuales temáticas.

Según el manual sobre *information gateways* publicado por el proyecto DESIRE⁸⁹, estas pasarelas de información temática son, frente a los buscadores y otros SRII, servicios de información basados en un control de calidad. Comentamos a continuación algunas de las características que concurren en todos ellos y que nos sirven para diferenciarlas por un lado, de los SRII de carácter global, y por otro de otras bibliotecas digitales productoras y poseedoras de su colección digital:

- a) Son de servicios en línea que proporcionan enlaces a otros sitios o documentos de Internet. Este rasgo no diferenciaría en principio, una *subject gateway* de cualquier otro sitio de la WWW, o de cualquier índice o directorio que base la búsqueda en la navegación.
- b) Su constitución se basa un proceso intelectual de selección de recursos conforme a la calidad de publicación y otros criterios propios del alcance del servicio (esto excluye, por ejemplo, la selección basada en la popularidad del sitio o de la página). Este es uno de los factores que diferencian estos servicios de motores o sistemas de recuperación automatizados de carácter global que, como Google, basan la selección de los recursos que indizan automáticamente en la técnica del *page-rank*⁹⁰.

⁸⁹ DESIRE. *Information Gateways...* Op. cit. <http://www.desire.org/handbook/print4.html>, [p. 1-2 en la versión impresa, o en el URL: <http://www.desire.org/handbook/print1.html#1-1>] (consultado el 1 de junio de 2000).

⁹⁰ Cfr. 7.2.

- c) Las descripciones del contenido son también de carácter intelectual, realizadas según un espectro de posibilidades de caracterización del contenido documental que van desde una pequeña anotación sobre el recurso, a una revisión exhaustiva del mismo (esto excluye los sumarios o resúmenes extraídos automáticamente). Una forma habitual, aunque no necesaria, de realizar análisis de contenido de los recursos seleccionados, es la asignación de descriptores o términos controlados.
- d) Las clasificaciones, índices o estructuras de navegación que presentan estos servicios se crean de forma manual, basándose para ello en estructuras lógicas de conocimiento (esto excluye completamente listas de *links* desestructuradas). Esta es otra característica de las *subject gateways* que las diferencia de otros sitios mal-llamados bibliotecas digitales, que no son más que listas de enlaces sin una definición de estructura intelectual previa, o con criterios de organización demasiado vagos para poder considerarlos bibliotecas.
- e) La característica fundamental, que acorde a los planteamientos de esta tesis, nos hace considerar a las *subject gateway* como bibliotecas digitales, es que en ellas se crean manualmente, al menos en parte, metadatos bibliográficos para los recursos seleccionados.
- f) La diferencia fundamental de estos servicios con respecto a otras bibliotecas productoras de DLOs es la formación de la colección. Aunque también podemos considerarlas como sistema, los subsistemas que forman una *subject gateway* (Fig. 61) son fundamentalmente el administrativo y el bibliográfico.
- El subsistema administrativo, además de incluir actividades relacionadas con la planificación del servicio y la evaluación tecnológica para fundamentarlo, conlleva el aspecto, nuclear en su concepción, de planificación de la selección de recursos electrónicos.
 - El subsistema bibliográfico, por su parte y desde nuestro punto de vista, conlleva una representación de la información post-publicación, y refiere a una colección virtual seleccionada de entre todos los recursos visibles a

través de un AU convencional o navegador en la WWW (publicados en HTML/XHTML⁹¹), que a su vez, se confunden con el universo de comunicación social global que fomenta y posibilita Internet. Es decir, los metadatos, que igualmente fundamentarán el sistema de consulta y de recuperación de la biblioteca, están creados con posterioridad a la creación y a la publicación del recurso en la WWW, y por agentes distintos a los autores/editores de los mismos.

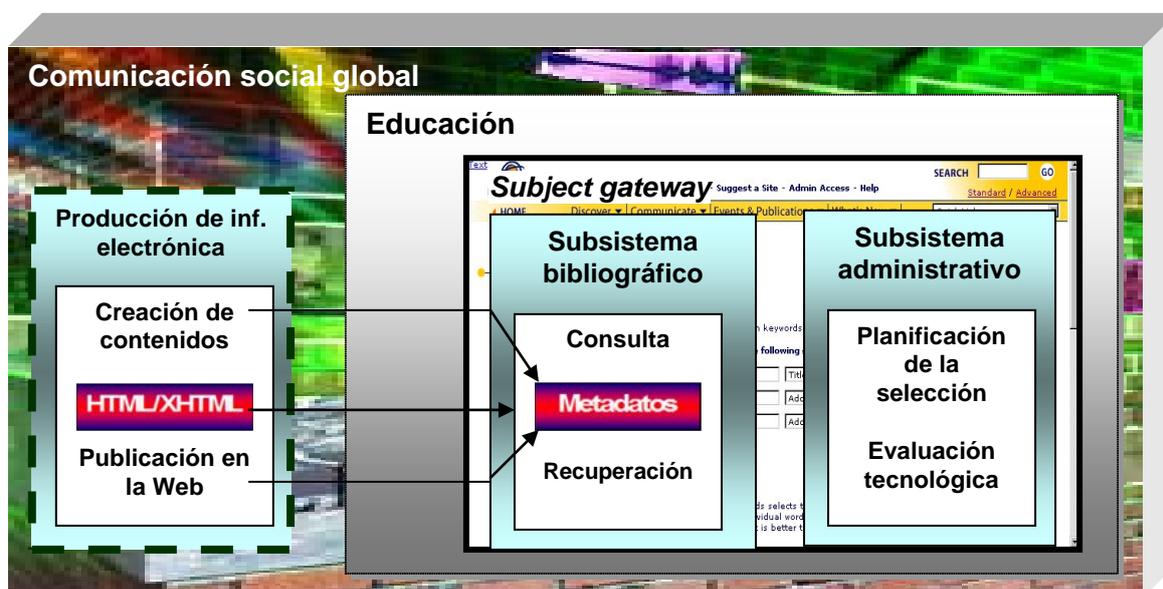


Fig. 61. Reformulación de la biblioteca como sistema en el mundo digital (Colecciones virtuales/Subject gateways)

- g) La producción de la información electrónica no forma parte de la constitución de estas bibliotecas digitales, es decir, no existe el subsistema bibliotecario de

⁹¹ Ello implica que los únicos metadatos que tendrán *a priori* los recursos Web seleccionados serán los que haya decidido incluir el autor del recurso, que estarán mediatizados por el software de edición de documentos Web que haya utilizado o por la aplicación específica de creación de metaetiquetas (Cfr. 4.3.1.). En cualquier caso, y en el mejor de ellos, la metainformación asociada a los recursos estará constituida por etiquetas <META> embebidas en el documento, normalmente <META NAME="keywords"> y <META NAME="description"> a lo sumo, la semántica del DC en sintaxis HTML, que podrían tomarse como referencia para la asignación posterior de metadatos propiamente dicha que se realizará, por terceros. *Vid. infr.*

edición o producción de información electrónica que señalábamos como parte integral de la concepción de la nueva biblioteca virtual como sistema.

La creación de contenidos, el marcado textual o la generación de otro tipo de DLOs, así como su publicación en la Web se realiza de forma totalmente ajena a la *subject gateway*. Esta circunstancia la aproxima más a las bibliotecas tradicionales que normalmente no crean los materiales de su colección, sin embargo también es la circunstancia que más las aleja de ellas en tanto que la colección es virtual, es decir, es un constructo no tangible, ni almacenado en la biblioteca, que se encuentra distribuido en la WWW.

- h) Por otra parte, este tipo de servicios vinculan nuevamente a la biblioteca digital, como concepto, a la base del proceso educativo, investigador y científico, que señalaba la obra clásica *La biblioteca: bibliosistemática e información*⁹²; ya que su origen y finalidad es, en muchos casos, complementar, de forma organizada e independiente, los servicios tradicionales que da una biblioteca en un entorno educativo.

Esta característica es lo que aproxima a las *subject gateways* al concepto de biblioteca híbrida, que trata de agrupar el servicio tradicional de información referencial a través de catálogo o de bases de datos bibliográficas, con un servicio de acceso al documento electrónico en texto completo. Esta dimensión de la biblioteca dentro de un marco educativo más amplio es lo que cimienta la idea de que el catálogo/s de las bibliotecas modernas (digitales y tradicionales) deben reflejar más que lo que la biblioteca posee, los recursos a los que ésta da acceso.

- i) En las *subject gateways* también confluyen las dos proposiciones básicas de la RI que señalábamos al final del capítulo anterior (Fig. 58): se basan en una colección con un límite definido (normalmente basado en la materia y en la calidad de los

⁹² Berta Enciso Carvajal. *Op. cit.*, <http://www.cervantesvirtual.com/servlet/SirveObras/079596188953519752324333>

recursos); y el usuario, aunque potencialmente es cualquier individuo conectado a Internet, normalmente es un usuario local, o al menos previsible por el carácter temático de estos servicios, que está familiarizado con la estructura de la información buscada.

Con todo, el objetivo de las *subject gateways* es mejorar el acceso a los recursos de información electrónica. La valoración de este tipo de bibliotecas digitales se puede enfocar —y así se trata en distintas investigaciones y proyectos sobre este tema— desde dos orientaciones:

- Desde el punto de vista de la selección, partiendo para ello de su función como organismos evaluadores de recursos Web, más que como servicios de información. Según esto, por ejemplo Codina traduce *information gateway* como "agencia de evaluación". Una agencia de evaluación, siguiendo a este autor, es el *organismo del sector público o privado que busca, selecciona y describe recursos digitales de acuerdo con un procedimiento explícito de selección y evaluación basado en la calidad del recurso*⁹³. Codina señala también que estas agencias hacen públicos los criterios de evaluación de los recursos que incluye y tienen como finalidad servir a los usuarios académicos y/o profesionales.
- Desde el punto de vista de la calidad del servicio. En este sentido se concibe o se trata: como sistema de recuperación de información basado en metadatos (que es el aspecto que más nos interesa en esta investigación), como un sistema de navegación basado en vocabularios controlados (tesauros y clasificaciones), y en general, se entienden como sistemas automatizados de acceso a la información electrónica por la importancia que se le concede al proceso de planificación tecnológica (tanto desde el punto de vista del software como del hardware, así

⁹³ Lluís Codina. Evaluación de recursos digitales... *Op. cit.*, p. 14.

como en el desarrollo de interfaces que respondan a una correcta accesibilidad y usabilidad⁹⁴).

A través de la observación de estos sistemas, podemos resumir que la aproximación que adoptan las bibliotecas virtuales temáticas con relación a los metadatos es la siguiente:

- Creación externa (*third party creator*⁹⁵, *post-publishing representation*⁹⁶) o perspectiva del tercer camino de la asignación de metadatos que señalaba Berners Lee⁹⁷.
- El responsable de la metainformación es un especialista o un profesional de la información (el gestor del depósito documental) que no tiene ninguna implicación con la creación ni con la edición del recurso digital.
- Implica la adopción y/o adaptación de un modelo del modelo de metadatos que mejor defina las características de la información (de los DLOs) que forman la colección.

Estas circunstancias son las que proporciona el control de calidad sobre el sistema de búsqueda, ya que las descripciones de los recursos (metadatos) se realizan de forma normalizada lo que dota, en principio, de coherencia semántica en todo el sistema de RI. La elección del formato de metadatos es una de las decisiones más importantes en la creación de *subject gateways*. La selección de un modelo

⁹⁴ Esta circunstancia se debe a que son servicios diseñados a por instituciones públicas y tratan de cumplir las pautas de accesibilidad y las recomendaciones de usabilidad para facilitar el acceso a todos los ciudadanos, en consonancia con las políticas de acceso. Vid. WCAG: <http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505>

⁹⁵ Lorcan Dempsey and Rachel Heery. Metadata: A Current View... *Op. cit.*, p. 165.

⁹⁶ Jian Quin. *Op. cit.*, p. 84.

⁹⁷ Tim Berners-Lee. Metadata Architecture... *Op. cit.*, <http://www.w3.org/DesignIssues/Metadata.html> Cfr. 4.2.

normalizado garantizará en mayor medida la compatibilidad con la aplicación informática para la de indización de metadatos (*harvester*⁹⁸) y la adaptación futura a la evolución de la metainformación.

Este modelo de biblioteca digital temática distribuida (*subject gateway*) tiene su origen el programa británico eLib⁹⁹ (*electronic Libraries Programme*) cuyo lema es: *la explotación de las tecnologías de información (TI) es esencial para crear el servicio bibliotecario del futuro*. eLib surgió a raíz de las recomendaciones del informe Follet (*Joint Funding Council's Libraries Review Group: Report*¹⁰⁰), realizado en diciembre de 1993, donde un grupo de expertos, encabezados por el profesor Sir Brian Follett e instados por los Consejos británicos de Financiación de la Educación Superior, evaluaron las bibliotecas de ese país y lanzaron como recomendación clave, la inversión en tecnologías de la información. Posteriormente, en 1995, se constituyó el Grupo Follet de Implantación de Tecnologías de la Información (FIGIT), en cuyo seno nacería el proyecto eLib, y a su vez diversas pasarelas de materia, como: SOSIG (*Social Science Information Gateway*), OMNI (*Organising Medical Networked Information*), EEVL (*Edinburgh Engineering Virtual Library*), ADAM (*Art, Design, Architecture and Media Information Gateway*), IHR-info (*Institute of Historical Research*), etc¹⁰¹. Para crear estas *subject gateways*, el programa eLib desarrolló, dentro del proyecto ROADS (*Resource*

⁹⁸ Cfr. Glosario, *harvest-harvesting /gather-gathering*.

⁹⁹ Para más información sobre eLib, Vid. <http://www.ukoln.ac.uk/services/elib> Cfr. Capítulo 3, notas 24 y 107 (WHOIS++).

¹⁰⁰ *Joint Funding Council's Libraries Review Group: Report (The Follet Report)* [documento HTML]. Bath: University, UKOLN, December 1993, rev. 19 de febrero de 1999. Disponible en: <http://www.ukoln.ac.uk/services/papers/follett/report> (consultado el 14 de diciembre de 1999).

¹⁰¹ Un listado completo de los proyectos desarrollados en el Reino Unido bajo el programa eLib, se puede consultar en: <http://www.ukoln.ac.uk/services/elib/projects> Además de estas pasarelas temáticas existen otros proyectos como bibliotecas híbridas, proyectos de conservación, publicaciones digitales, así como múltiples investigaciones relacionadas con la recuperación de información distribuida, como por ejemplo MODELS (*MOVing to Distributed Environments for Library Services*).

Organisation And Discovery in Subject-based Services), un conjunto de herramientas de software que permiten la construcción de servicios de indización y búsqueda distribuidos. ROADS es por tanto, el paquete de software que permitió crear y mantener una base de datos accesible vía Web, de registros de metadatos relevantes para una materia particular (medicina, química, arquitectura, etc.) en el contexto de la información académica inglesa.

Aunque todas estas *gateways* temáticas, tienen las particularidades de los recursos electrónicos sobre la materia que tratan, comparten características comunes, según Dempsey¹⁰²: la finalidad es facilitar servicios de búsqueda en Internet a las organizaciones de educación superior del Reino Unido; se basan en formatos de metadatos situados en la banda 2, fundamentalmente el DC; están desarrollando bases de datos fundamentadas en la descripción de recursos de los que se derivan otros servicios; tienen definidos unos criterios de selección y calidad que determinan qué recursos se incluyen en el sistema; son sistemas de búsqueda y recuperación centrados en el ámbito servidor, pero aspiran a cambiar a un nivel de objeto; generalmente el esfuerzo del personal que trabaja en la construcción de estos sistemas está dirigido a la gestión, administración, catalogación, desarrollo técnico y promoción.

Este modelo de biblioteca digital distribuida ha tenido mucho éxito en los tres últimos años (1998-2001) en el resto del mundo. Por ejemplo en Australia¹⁰³, el sector de la educación superior ha desarrollado iniciativas parecidas a las británicas centrándose en tres disciplinas (agricultura, química e ingeniería) y dando lugar a las

¹⁰² Lorcan Dempsey. Roads to Desire: Some UK and Other European Metadata and Resource Discovery Projects [documento HTML]. *D-Lib Magazine*, July/August 1996. Disponible en: <http://www.dlib.org/dlib/july96/07dempsey.html> (consultado el 18 de junio de 1998).

¹⁰³ Para una completa visión de las pasarelas temáticas australianas, sus características, peculiaridades y perspectiva, *Vid.* Debbie Campbell. Australian Subject Gateways: Metadata as an Agent of Change [documento PDF]. En: *VALA Biennial Conference and Exhibition (10. 2000. Melbourne)*. Disponible en: <http://www.vala.org.au/vala2000/2000pdf/Campbell.PDF> (consultado el 18 de marzo de 2001).

respectivas pasarelas temáticas: AGRIGATE (*Agriculture Information Gateway for Australian Researchers*) en la Universidad de Melbourne, MetaChem en la University of New South Wales y AVEL en la Universidad de Queensland junto con proyectos como EdNA (*Educational Network Australia*) o la *Australian Digital Thesis* (ADT). En Brasil, el *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) también ha seguido este modelo creando las *bibliotecas virtuais tematicas* <<http://www.prossiga.br/bvtematicas>> sobre distintas materias (incluso, *biblioteca virtual sobre bibliotecas virtuais* <<http://www.prossiga.br/bibvirtual>>), que tratan de emular a las *subject gateways* británicas.

Los aspectos clave de estos sistemas de información basados en metadatos, son dos:

- 1) La distinción entre servidores, que permiten el almacenamiento de los objetos (DLOs) que producen, y el concepto de colección distribuida que supone la selección y organización de dichos documentos entendidos como objetos.
- 2) Introduce el concepto de recuperación de información distribuida: basado en un modelo donde varios servidores independientes, indizan colecciones locales de documentos y un servidor directorio, guía a los usuarios a los índices independientes. El modelo de indización distribuida descrito por Sheldon¹⁰⁴ y estudiado por gran cantidad de autores a tenor de la interoperabilidad necesaria entre sistemas heterogéneos de codificación de información¹⁰⁵, tiene la característica de describir una colección a través de una etiqueta de contenido, donde tal etiqueta puede procesarse por sí misma como un documento e incluirse,

¹⁰⁴ M. Sheldon. *Content Routing: A Scalable Architecture for Network-Based Information Discovery*. PhD thesis, MIT Dept. of EECS, Oct. 1995. Citado en: Jeremy Hylton. Corp. For National Research Initiatives Position Paper: Creating collection with a distributed indexing infrastructure. En: *Distributed Indexing... Op. cit.*, <http://www.w3.org/Search/9605-Indexing-Workshop/Papers/AllPaps.doc>, p. 21.

¹⁰⁵ Cfr. *infr.* 8.3.2.

a su vez, en otra colección. Según este autor, las etiquetas de contenido ayudan a los usuarios a gestionar y examinar espacios de información muy amplios, pero la idea puede extenderse gracias al procesamiento de colecciones (y sus metadatos asociados) de forma independiente de sus servidores, de tal manera que una colección pueda incluir documentos particulares de distintos servidores. Como comenta Denenberg¹⁰⁶, los objetos en una colección pueden distribuirse, pero la colección se presenta cohesionada lógicamente, porque las estructuras de metadatos están asociadas a ella y soportan una navegación coherente

Estas bibliotecas virtuales temáticas, organizadas con sistemas de metadatos, constituyen el servicio de información de carácter semántico que protagoniza una facción de la realidad de la aplicabilidad de los metadatos en la actualidad. Se basa en la imposibilidad de controlar los procesos de edición de información en un entorno universal y abierto como Internet y constituyen un estadio evolutivo hacia la Web Semántica, ya que homogenizan (aunque sea *a posteriori*), según formatos normalizados de estructuración de metainformación, conjuntos o colecciones de DLOs.

Tanto estas colecciones descritas y estructuradas *a posteriori*, como las bibliotecas digitales productoras de sus documentos que estructuran su metainformación *a priori*, deben utilizar estándares de metadatos de tal forma que puedan interoperar (y que esa es su tendencia) en la creación de un sistema de información semántico y globalizado. Todos los sistemas que hemos analizado en este apartado (bibliotecas digitales editoras de su fondo digital y *subject gateways* que dan acceso a colecciones virtuales distribuidas), así como otras realidades informativas de características similares que no mencionamos ya que su análisis rebasaría los límites de esta investigación, son los servicios que, dentro de la miríada de información de la Red y partiendo de nuestra concepción de los metadatos, podemos considerar bibliotecas digitales. Bibliotecas espurias de sus homólogos

¹⁰⁶ Ray Denenberg. *Op. cit.*, <http://lcweb.loc.gov/z3950/agency/papers/italy.html>

tradicionales que tratan de emularlas constituyendo una infraestructura de sistemas de metadatos, sistemas de indización y búsqueda, en los que confiamos para fundar una recuperación de información más coherente, más eficaz, que se convierta en una Web más semántica.

8.3. Bibliotecas digitales y globalización

Como afirman Terceiro y Matías¹⁰⁷, *la convergencia tecnológica es la causa más importante de la globalización*, lo que correlativamente implica que la globalización es consecuencia de la convergencia tecnológica. También podríamos pensar a la inversa, que la convergencia tecnológica es consecuencia de la globalización, o mejor aún que globalización y convergencia tecnológica tienen una causalidad recíproca, al menos si partimos, como en esta tesis, de otros planteamientos lejanos a la tendencia neoliberal del mercado. En cualquier caso, las tecnologías de la información y la globalización están íntimamente relacionadas entre sí, y ambas a su vez, están vinculadas a la construcción de un nuevo paradigma tecno-económico o tecno-social, que se basa en la tecnología, que contribuye a entender el nuevo desarrollo de la sociedad con respecto a las TICs y a la propia información. El deseo de proporcionar un acceso mundial y universal a la información requiere también un nuevo planteamiento en las bibliotecas, un cambio también en la gestión de la información.

La difusión masiva de las tecnologías tiene como consecuencia, además de esa divulgada, ansiada y también odiada globalización, cambios radicales en las instituciones públicas y privadas así como la necesidad de nuevas políticas de información. El replanteamiento político a tenor de la Sociedad de la Información conlleva, entre otras características: una publicación masiva a través de la WWW que se convierte en un medio primigenio para el intercambio y difusión de la

¹⁰⁷ José B. Terceiro y Gustavo Matías. *Op. cit.*, p. 79.

información, la creación de nuevos mercados y el desarrollo del comercio electrónico, la creación de servicios de teleeducación, la expansión de la investigación a través del desarrollo de comunidades, la participación de individuos y grupos en los procesos políticos, y sobre todo, en el ámbito que concierne a esta tesis, el desarrollo intensivo de contenidos en formato electrónico y por ende, el desarrollo de unidades de información (bibliotecas y archivos) digitales¹⁰⁸.

La globalización de la información no se puede entender sin hacer referencia a la propia lógica de la mundialización de la economía, de la política y de la cultura, que conllevan problemas y retos de distinta naturaleza. El replanteamiento profesional alcanza incluso a aspectos doctrinales en determinadas disciplinas, y sin duda, a los principios y las prácticas de la gestión de la información. Los metadatos, como hemos dejado patente en nuestra tesis, son la clave de la redefinición de la práctica bibliotecaria, que constituirán, de manera normalizada, el principal cimiento de la organización y recuperación de la información a un nivel mundial. Sin embargo, como afirman Kashyap y Sheth¹⁰⁹, *la tendencia de globalización en los sistemas de información ha incrementado la complejidad de la sobrecarga de información a través de la heterogeneidad de los datos digitales y de los metadatos asociados a ellos para su recuperación.*

Aunque hasta ahora hemos centrado nuestra tesis en la validez, utilidad y operatividad de los metadatos en entornos controlables, "finitos", locales o temáticos,

¹⁰⁸ Así se reconoce por ejemplo, en el planteamiento del V Programa Marco de la Unión Europea para la Investigación y desarrollo tecnológico que incluye entre sus acciones clave (dentro del programa IST para hacer una *sociedad de la información fácil de usar*), el desarrollo de herramientas y contenidos multimedia, destacando, *la publicación interactiva con nuevos métodos para crear y estructurar publicaciones; la diseminación personalizada de la información y el acceso al conocimiento científico y cultural a través de bibliotecas, archivos y museos en red. Decision N° 182/1999/EC of the European Parliament and of the Council of 22 December 1998 Concerning the Fifth Framework Programme of the European Community for Research, Technological Development and Demonstration Activities (1998 to 2002)* [documento HTML]. Cordis, 1999. Disponible en: <http://www.cordis.lu/fp5/src/ec-en5.htm> (consultado el 25 de mayo de 2000).

¹⁰⁹ Virpul Kashyap, Amit Sheth. *Op. cit.*, p. 4.

de información, avalando para ello nuestras proposiciones en la realidad tecnológica actual, la tesis quedaría incompleta si no reflexionáramos sobre la vocación global de Internet y cómo los metadatos pueden contribuir a reactivar la utopía de una biblioteca universal. Para ello, en este apartado trataremos el contexto del acceso universal a la información, las implicaciones que esto conlleva en términos de interoperabilidad, para soslayar el problema de la heterogeneidad de los modelos biblioteca digital y de los estándares de metadatos asociados a ellas.

8.3.1. El acceso universal a la información

El marco estructural de esta investigación deja patente los cambios que ha promovido la Sociedad de la Información, donde el futuro de las bibliotecas, fruto de la contemporaneidad de la relación que establece el hombre con la tecnología, es inminentemente digital. Estamos pues en la Era Web, como definíamos en la primera parte de este capítulo y el ritmo de cambio en la sociedad humana continúa acelerándose. Las TICs son las impulsoras de ese cambio y el capital intelectual se ha convertido en el primer factor de desarrollo en una sociedad protagonizada por la información. El acceso eficaz a la información a través de la tecnología se ha convertido en el centro de las políticas públicas internacionales de información, surgiendo el nuevo concepto de *acceso universal* al conocimiento.

El acceso a la información y al conocimiento se ha convertido en un factor esencial del desarrollo de tal forma que, tanto los gobiernos internacionales, a través de la política de información del G7-G8, como instituciones de carácter privado como la GKP¹¹⁰ (*Global Knowledge Partnership*) afirman que el acceso y uso eficaz de las TICs puede mejorar el crecimiento económico y el bienestar de todos los ciudadanos. Ese acceso universal a la información implica que existe una disponibilidad global de estrategias y herramientas para el uso de la información, esto

¹¹⁰ Vid. Global Knowledge Partnership: <http://www.globalknowledge.org>

es, el acceso tanto a infraestructuras y servicios como a infoestructuras y contenidos. Así pues, gobiernos e instituciones públicas y privadas están impulsando, a distinto nivel, el desarrollo de la Sociedad de la Información para potenciar el acceso universal y la participación de todos los ciudadanos en la economía global del conocimiento. En el nuevo milenio, el acceso a la información es un derecho fundamental, asimilado e instigado al acceso a la tecnología.

Todos estos cambios socio-políticos implican que la calidad de las decisiones depende de la calidad de la información y que el acceso a la información y al conocimiento resulta en una mejora de la calidad de vida, siempre que exista un contexto social que lo posibilite. El reto para los sistemas de información y bibliotecas, ante esta nueva realidad, implica la investigación en nuevas formas de facilitar ese acceso universal.

Esta realidad estructural que impera en el desarrollo de servicios y sistemas de información para la nueva Sociedad del Conocimiento, se traduce, en el contexto de este trabajo en que: los modelos de metadatos y la formalización de estándares y prácticas en torno a ellos, así como la constitución de bibliotecas digitales, implican un planteamiento abarcable de organización de información —normalmente de carácter temático—, local a pesar de su proyección global, que optimiza el acceso a una información de calidad parcializada a la tipología de DLOs o a su temática de contenido. Sin embargo, en el contexto actual de la globalización de la información, la necesidad de llegar a un usuario global, provoca la improbabilidad en la persistencia de un formato vinculado a un solo grupo de interés o a un colectivo profesional (comunidades virtuales o CoIs) y la necesidad de crear nuevas formas de interrelación de esos servicios finitos y abarcables para los que plantemos los metadatos.

Según se afirma en la ARL (*American Research Libraries*¹¹¹), el acceso universal a los servicios de información es uno de los objetivos principales de la biblioteca digital. Las colecciones no se limitan a sustitutos (*subrogates*) de los documentos, se extienden también a objetos digitales que no pueden ser representados o distribuidos en formato impreso y que su forma original es electrónica. El acceso global de las bibliotecas digitales hace que la localización física del almacenamiento de la información sea menos importante que en las bibliotecas tradicionales basadas en materiales impresos. Sin embargo, la idiosincrasia local y regional de los usuarios deberá adaptarse, aunque sea lentamente, a una visión global, sobre todo cuando las bibliotecas digitales trasciendan los dominios estrictos de la investigación y de las grandes organizaciones.

La globalización es un factor inminente y fundamental en todos los ámbitos sociales; en lo que respecta a las bibliotecas y la organización de la información electrónica, conlleva implicaciones de índole diversa como el multilingüismo o la preservación de tradiciones culturales y patrimoniales diferentes. En este sentido podemos conferir a los metadatos una responsabilidad tácita especial de carácter cultural, nos referimos a la necesidad de reconciliar los sistemas de información locales con la accesibilidad global. El papel de las bibliotecas de proporcionar la descripción y el acceso universal a los recursos de información estará vinculado a la creación, mantenimiento y escalabilidad de las descripciones de metadatos. Si trasladamos toda la discusión de los metadatos, de la que venimos dando cuenta a lo largo de todo este trabajo, al nuevo contexto bibliotecario protagonizado por la información electrónica, es evidente que el futuro del acceso universal no estará aparejado a un único sistema mundial de descripción, sino a la interoperabilidad entre

¹¹¹ Citado en: Sarah B. Watstein, Pascal V. Calarco, James S. Ghaphery. *Op. cit.*, p. 349.

formatos. Lagonze¹¹² señala algunas de las características a las que debe enfrentarse la gestión de la información electrónica a partir de metadatos, ante el paradigma del acceso universal a la información:

- *Especialización*. Los formatos y esquemas suelen reflejar las necesidades de las comunidades específicas que los crean.
- *Descentralización*. La producción y el mantenimiento de metadatos surge en el seno de distintas comunidades científicas, con distinto nivel de experiencia, que raras veces comparten prácticas comunes o normas.
- *Democratización*. Algunas iniciativas de metainformación, sobre todo la DCMI, como hemos mencionado¹¹³, están dirigidas a un colectivo no profesional, de tal forma que los metadatos puedan crearse y mantenerse por grupos no especializados de usuarios.

La globalización en torno al mundo de la información pone en tela de juicio la persistencia de formatos de codificación de información propios o exclusivos de un área del conocimiento. Esta realidad implica la necesidad de adaptar los formatos de procesamiento e intercambio de la información al nuevo contexto. Así, en el tratamiento de la documentación electrónica surgen formatos de intercambio y estándares genéricos como los que resaltamos en el capítulo 5 (SGML/XML), surge también la necesidad de describir claramente y sin ambigüedades los diferentes elementos (lógicos o conceptuales) de los documentos, la integración de elementos

¹¹² Carl Lagoze. Business Unusual: How "Event-Awareness May Breathe Life into Catalog? [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 19 de octubre de 2000, rev. 16 de noviembre de 2000. Disponible en: http://www.unicamp.br/bc/lagonze_paper.html (consultado el 21 de febrero de 2001).

¹¹³ Cfr. 4.2. y 5.2.1.2

de acceso dentro del propio documento (como las DTDs) y la gestión de la integración de distintos DLOs (texto, imágenes, vídeo, audio, etc.).

Debemos mencionar un proyecto piloto, que con el nombre de *Bibliotheca Universalis*¹¹⁴ recoge toda la filosofía que debe infundir la gestión de la información en bibliotecas digitales cuya proyección debe de ser global. Surgido a raíz de los principios emanados de la reunión ministerial del G7 celebrada en 1995 en Bruselas para el desarrollo de una Sociedad de la Información Global, tiene como objetivo poner a disposición de los ciudadanos, a través de las redes, la amplia cantidad de conocimientos derivados de los actuales programas de digitalización. Según la evaluación provisional de este proyecto, realizada por el G8, *Bibliotheca Universalis* está consolidando la cooperación internacional para el establecimiento de sistemas de "bibliotecas electrónicas" mundiales. Este proyecto ayudará a las bibliotecas a mejorar la disponibilidad en el ámbito internacional de objetos de información digitalizados y a promover la digitalización de un gran volumen de materiales informativos. Con posterioridad al informe provisional del G8, en la conferencia de la IFLA en Ámsterdam de 1998, se firmó un acuerdo formal con bibliotecas nacionales y otras instituciones sobre la cooperación en este proyecto, quedando abierto para todas aquellas bibliotecas nacionales que estuvieran digitalizando sus colecciones. La biblioteca Nacional de Canadá, junto al grupo de trabajo de CoBRA+ (*Computerized Bibliographic Record Actions*), identificarían los problemas técnicos relativos a los estándares y métodos de acceso y propondrían recomendaciones para garantizar la interoperabilidad entre estas bibliotecas digitales¹¹⁵.

¹¹⁴ *Bibliotheca Universalis: a G-7 Global Information Society Pilot Project* [documento HTML]. London: British Library, Gabriel, 1998. Disponible en: <http://www.bl.uk/gabriel/bibliotheca-universalis/index.html> (consultado el 18 de mayo de 1999).

¹¹⁵ *Bibliotheca Universalis Project: Partners, Objectives and Achievements 1996-1998* [documento HTML]. London, etc.: Gateway to Europe's National Libraries, rev. 29 de septiembre de 1998. Disponible en: <http://portico.bl.uk/gabriel/bibliotheca-universalis/bibuniv.htm> (consultado el 7 de junio de 1999).

Para incluir información bibliográfica en las bibliotecas digitales, los metadatos se expresan normalmente siguiendo estándares de metainformación como el DC o el MARC. Sin embargo, es necesario generalizar la descripción de datos y metadatos que hagan accesibles, de manera coordinada, una amplia gama de bibliotecas digitales. La aceptación generalizada de XML será un factor decisivo en este sentido. Las etiquetas del documento pueden usarse para describir el significado de los componentes de un DLO. Por ello, controlar la semántica asociada al marcado XML será una tarea decisiva. Esta circunstancia abre una nueva perspectiva en el acceso a las bibliotecas digitales, en tanto que XML se ha convertido en el estándar *de facto* para codificar la información en las bibliotecas digitales, fundamentalmente en aquellas que son responsables de la edición de sus fondos digitalizados, pero también en las que basan su colección en recursos distribuidos por toda la WWW.

Con todo lo que hemos comentado a lo largo de esta tesis, podemos prever una situación futura donde XML puede utilizarse tanto para el intercambio de documentos digitales, incluso multimedia (DLOs), XSL para la presentación multimodal (HTML-XML, etc.) y para la definición de metainformación, la utilización de *schemas* o DTDs XML a los que se asocien esquemas RDF para la descripción. Incluso desde una perspectiva de los metadatos orientados a la materia, se puede prever una aproximación basada en ontologías que intenten construir una ontología común para las bibliotecas digitales, de tal forma que representen una visión global "virtual" de todas ellas y definan las reglas para convertir sus visiones de contenido local, en visiones de contenido global. Esta situación futura, es lo que Tim Berners Lee denomina Web Semántica y que permitirá la organización de bibliotecas digitales globales.

8.3.2. Metadatos e interoperabilidad: la necesidad de integración

Como vimos en el apartado anterior (8.2.), Internet está promoviendo el acceso a un número amplio, diverso y creciente de sistemas de información digital. Bibliotecas digitales originalmente concebidas como sistemas "locales" de

recuperación de información para un grupo especializado de usuarios (entendido éste en algunos casos, comunidades virtuales unidas por intereses comunes, un mismo idioma, etc.) pero cuya proyección, gracias a la universalidad de la WWW, alcanza a una amplia gama de usuarios potenciales y un acceso previsiblemente global. La constitución de bibliotecas digitales implica la integración de complejos sistemas que incluyen colecciones de documentos con estructuras diferentes, distintos medios y contenidos de diversos tipos. Junto a esto tenemos que considerar una mezcla sugestiva y variada de componentes de hardware y software, que pretenden interoperar a través de diferentes estructuras de datos, algoritmos de procesamiento y múltiples personas, comunidades e instituciones que tienen objetivos, políticas y culturas diferentes. El reto de la integración y de intercambio de información, está sobre todo vinculado a grupos del ámbito académico, gubernamental o/y corporativos de temática específica.

En la segunda parte de este trabajo trazamos una visión de algunos de los más importantes modelos de metadatos, donde translucen sus aspectos de confluencia o sus diferencias de cobertura más patentes, así como su prospectiva de normalización. También hemos comentado que son muchos los tipos de metadatos que existen según el ámbito de aplicación o dominio de información en que se utilizan. Todos estos formatos normalmente reflejan la naturaleza de los objetos de información que describen y los objetivos de la comunidad a la que sirven; así son diferentes, p. ej., los metadatos creados para un recurso gubernamental (según el formato GILS) de los utilizados para una imagen de una pintura renacentista (que base su descripción en el modelo VRA), dando todo ello lugar a una gran disensión semántica de las etiquetas de distintos formatos y por tanto también a una diversidad de sistemas de recuperación (locales aunque de proyección global) fundados en ellos.

La existencia de esa gran variedad de modelos representa el intento de reunir las necesidades de información que tienen las diferentes comunidades científicas. Estos colectivos implicados en la descripción de recursos han realizado esfuerzos importantes por desarrollar estructuras especializadas, incluidas dentro de sus propios sistemas de información, que les permitan enriquecer la descripción de registros de

tal forma que respondan a sus exigencias concretas. Como ya hemos comentado, es improbable que se use un formato de metadatos monolítico para la descripción y recuperación de recursos en la compleja infoestructura de la Sociedad de la información. A esto se debe añadir también que es improbable que se use de manera universal aunque se normalice formalmente, porque es difícil que un sólo formato se adecue a todas las necesidades de todos los grupos o comunidades de información. Estos grupos o colectivos involucrados en la creación y el uso de diferentes bibliotecas digitales están sujetos a sus propios esquemas, lo que es bastante explicable si se toma en cuenta que han realizado un esfuerzo por conseguir un consenso para establecerlos, un nivel de destreza para aplicarlos de forma consistente y una gran inversión en todo ello.

El carácter abierto de Internet posibilita un entorno informativo donde acceder a cualquier tipo de datos es potencialmente factible. Pero si tenemos en cuenta que además que este entorno es heterogéneo y los distintos sistemas autónomos de información representan y estructuran su información de manera inconsistente, sumando la utopía de un formato único, es inevitable pensar en mecanismos de interacción entre los distintos sistemas de representación de datos. Ese mecanismo para recuperar los diversos recursos de información de la Web, es la creación de *crosswalks*¹¹⁶ que resulten del análisis y mapeo de las semejanzas de cada sistema de metadatos. Cromwell-Kessler¹¹⁷ opina, refiriéndose a la creación de *crosswalks*, que:

La creación de correspondencias entre los distintos sistemas de metadatos es el modo más eficaz de reconciliar sus diferencias y de facilitar la interoperabilidad entre sistemas para llegar así al acceso integrado. Se crea

¹¹⁶ Utilizamos el término en inglés *crosswalks*, porque no existe una traducción posible de esta palabra en el contexto en que la estamos utilizando. Vid. Glosario, *crosswalks/crossroads*.

¹¹⁷ Willy Cromwell-Kessler. Correspondencias entre metadatos e interoperabilidad: ¿qué significa esto? En: *Introducción a los Metadatos: vías...* Op. cit., p. 21. Cfr. versión electrónica actualizada en la contribución de Mary Woodley. Crosswalks: the Path to Universal Access?. En: *Introduction to metadata: pathways...* Op. cit., http://www.getty.edu/research/institute/standards/intrometadata/2_articles/woodley

de esta forma, la posibilidad de interacción entre los sistemas automatizados, haciendo posible el objetivo de un acceso integrado y global (como sucedió, por ejemplo, entre los distintos formatos MARC).

El primer obstáculo para esa integración en el acceso a sistemas distribuidos de información es la propia estructura plural de la metainformación, que implica la utilización de diferentes elementos de metadatos que funcionan a un nivel distinto. Podría darse el caso de que varias comunidades de usuarios se reunieran para reconciliar sus sistemas y desarrollar un estándar interdisciplinar pero, como apuntábamos antes, a pesar de que sea una posibilidad, es improbable que se origine esta solución. Y en el caso de que surgiera, habría que hacer una conversión retrospectiva de toda la información analizada según el primer modelo de metadatos (por ejemplo MARC) al nuevo estándar interdisciplinar de codificación de metainformación (XML) lo cual es desaconsejable en tiempo y coste. Por ello, la integración se producirá a través del traslado o traducción de un sistema de estructuración de datos a otro y viceversa. En este sentido los *crosswalks* de metadatos funcionan como una especie de clave o código *de facto* para traducir automáticamente de un formato a otro, permitiendo interoperar a ambos sistemas.

De todo lo que hemos dicho hasta ahora, dada la plétora de esquemas, modelos o formatos, se deduce que: la utilidad de un conjunto de metadatos estructurados para la recuperación de información más allá de subconjuntos de información o bibliotecas digitales, depende exclusivamente de la capacidad que tengan dichos formatos para establecer interoperabilidad entre sistemas y de que éstos interactúen en pos de una recuperación de información global. Si bien es cierto que ninguna biblioteca digital o sistema de información que tenga como base de su funcionamiento Internet, nunca puede entenderse como un sistema local, en el sentido de que funcionará en cualquier ordenador independiente del hardware y del software, aquí utilizamos la palabra "local" para referirnos a los entornos "finitos" y abarcables de información, perfectamente definidos, a los que nos referimos en esta tesis como entornos operativos para la gestión de la información a través de metadatos. Las bibliotecas digitales hoy en día tienen una proyección de alcance

universal, pero su funcionamiento de recuperación es local. Es necesario sin embargo que se vuelva a la concepción de un acceso global y un sistema de recuperación global, independientemente de su planteamiento de gestión de metadatos¹¹⁸.

Llegamos así, a la noción fundamental de este apartado, y al concepto clave también de la recuperación de información basada en metadatos: interoperabilidad. La interoperabilidad puede definirse, en el marco de los metadatos, como *la habilidad de dos o más sistemas o componentes para intercambiar información y utilizar dicha información intercambiada sin un esfuerzo especial en su sistema*¹¹⁹. Esa interoperabilidad requiere sobre todo, convenciones en torno a los sistemas de metadatos, desde el punto de vista la sintaxis (la combinación de las palabras o las gramáticas para comunicar semántica y estructura), de la estructura (que se pueda leer por el hombre y que puedan validarse a través de la máquina), del vocabulario (la correspondencia entre el símbolo y el concepto) y de la semántica (el significado de los elementos o del vocabulario).

¹¹⁸ Utilizamos aquí "global" por oposición a "local", como hemos hecho en otras partes de esta investigación. Hasta ahora en esta tesis, sólo nos habíamos referido a la recuperación de información global, al hablar de los sistemas de recuperación de la Web (motores, directorios, metabuscadores, etc.) evocando, en ese caso, a su carácter genérico y a la pretensión de dar acceso a "todos" los documentos de la WWW (*Cfr.* Capítulo 7). En este capítulo utilizamos "global", con una doble pretensión: por un lado, aludir a la proyección que tienen las bibliotecas digitales en la Era Web, y por otro reflejar la necesidad de acceso integrado entre todas ellas.

De esta forma podemos apuntar un efecto cíclico la parte III de esta investigación, marcado por la propia naturaleza del complejo mundo de la RI en la Red: primeramente hablamos de la imposibilidad de los SRII de asumir, a pesar de su cobertura global, un acceso de calidad a los contenidos de la WWW, donde la perspectiva del uso de un modelo único de metaetiquetas fracasa por la propia inconsistencia de su aplicación; después apuntamos la amplia realidad de las bibliotecas digitales que surgen para paliar la falta de calidad en la recuperación de información, donde se pierde la perspectiva de globalidad en pos de la calidad, desarrollando para ello, modelos de metadatos y sistemas de recuperación específicos, operativos para sus entornos finitos de información; y ahora introducimos la perspectiva de la cobertura global basada en la sinergia de esas soluciones locales de calidad, que se logrará a través de la interoperabilidad entre distintos sistemas y esquemas de metainformación.

¹¹⁹ American Library Association. *Committee on Cataloging: Description and Access. Task Force on Metadata: Summary Report June 1999* [documento HTML]. ALA, junio 1999, rev. 7 de agosto de 1999. Disponible en: <http://www.ala.org/alcts/organizations/ccs/ccda/tf-meta3.html> (consultado el 2 de septiembre de 1999). *Cfr.* Glosario, *interoperability*.

En el entorno de la investigación en RI, sobre todo en el vinculado a las bibliotecas digitales, muchas de las discusiones sobre infraestructura se centran en establecer herramientas comunes, tecnologías e incluso estándares compartidos, que puedan facilitar la base para estudiar distintos aspectos de la interoperabilidad. Sin embargo, en abstracto resulta difícil definir la interoperabilidad o, al menos, la reflexión que podemos hacer aquí tiene un valor limitado, ya que consideramos que para argumentar de forma consistente el concepto e implicaciones de la interoperabilidad en entornos digitales, requeriría una amplia experimentación con distintos sistemas que utilizan metadatos como base de la recuperación, lo que excedería nuestros objetivos. A pesar de esto, plantearemos una aproximación teórica a dicha noción que sirva, cuanto menos, para evidenciar su importancia.

La interoperabilidad supone complementar una serie de estrategias humanas, abocadas a la descripción consistente y flexible de la información mediante metadatos, con un conjunto de herramientas, software y protocolos que permitan que los distintos sistemas de información (bibliotecas digitales, fundamentalmente) puedan establecer *crosswalks* o mecanismos de traslado automático entre ellos para reconciliar sus diferencias y alcanzar, o al menos acercarse, a un acceso universal o integrado.

Lynch y García-Molina¹²⁰, identifican un espectro completo de puntos de vista de la interoperabilidad, que puede ser una buena forma de acercarnos a la cobertura conceptual del término y a sus implicaciones. A un lado de este espectro estaría, según estos autores, el uso de herramientas e interfaces comunes que proporcionan una uniformidad superficial al acceso y navegación, pero que se confía, casi por completo, a la inteligencia humana para dar alguna coherencia al contenido. Al otro extremo de ese espectro sitúan una interoperabilidad semántica amplia como uno de

¹²⁰ Clifford Lynch and Héctor García-Molina. Interoperability, Scaling and Digital Libraries [documento HTML]. En: *Research Agenda. IITA Digital Libraries Workshop, August 22, 1995*. Stanford: University, 22 de agosto de 1995. Disponible en: <http://www-diglib.stanford.edu/diglib/pub/reports/itta-dlw/main.html> (consultado el 24 de julio de 1998).

los retos más importantes en la interacción entre distintas bibliotecas o colecciones digitales. Esa interoperabilidad semántica, trata de ocuparse de la aptitud de los usuarios para acceder, de manera coherente, a similares (aunque definidos y gestionados de forma autónoma) tipos de objetos y servicios digitales distribuidos a través de depósitos documentales heterogéneos, con un software que compense sus divergencias. Finalmente, como término medio ente estos dos extremos (el que confía la interoperabilidad a la inteligencia humana y el que se relaciona con el software capaz de mantener una interoperabilidad semántica amplia), señalan la interoperabilidad sintáctica, fundamentada en el intercambio de metadatos y en el uso de protocolos de transmisión de objetos digitales y formatos basados en estos metadatos, como manera de proporcionar una coherencia limitada del contenido, completada por la interpretación humana.

Renato Iannella¹²¹, más axiomático, señala que la interoperabilidad se debe dar en cuatro niveles:

- *Semántico*: La interoperabilidad semántica es la posibilidad de entender cada elemento dentro de un esquema de metadatos. Por ejemplo en el DC, el elemento `Creator` se define como *la persona u organización responsable de la creación del contenido intelectual de un recurso*¹²². Sin embargo en el formato VRA se define este mismo elemento como *los nombres, denominaciones u otros identificadores asignados a una persona, grupo de personas, entidad corporativa, u otra entidad que han contribuido al diseño, creación, producción, manufactura o modificación del trabajo*¹²³. Por su parte, en la TEIH se define el

¹²¹ Renato Iannella. Mostly Metadata... *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA1998>

¹²² [Javier Massa Marín]. *Elementos del conjunto de metadatos [sic.] de Dublin Core: Descripción de referencia* [documento HTML]. RedIris, rev. 13 de mayo de 1998. Disponible en: http://www.rediris.es/metadata/dublin_core_elements.es.html (consultado el 17 de julio de 1998).

¹²³ *VRA Core Categories, version 3.0 Op. cit.*, <http://www.gsd.harvard.edu/%7Estaffaw3/vra/vracore3.htm>

elemento <AUTHOR> en los siguientes términos: *contiene el nombre del autor o de los autores, ya sean personales o corporativos de una obra; constituye la mención de responsabilidad principal para toda una unidad bibliográfica*¹²⁴. En estas definiciones se podría argumentar que los tres elementos según los diferentes formatos, son semánticamente equivalentes¹²⁵. Otro método de interoperabilidad semántica es definir un conjunto de elementos esenciales que son comunes a los diferentes esquemas. Iannella opina que la semántica es uno de los aspectos más difíciles de la interoperabilidad y que requiere un consenso dentro de las distintas comunidades de información que trabajan con metadatos.

- *Sintáctico*: La interoperabilidad sintáctica concierne a cómo se codifican digitalmente los metadatos. La sintaxis de cualquier modelo de metadatos, como hemos visto, refleja una forma estandarizada de generar y transmitir los metadatos, definiendo reglas para codificar los nombres de los elementos (tipos de propiedad en RDF) y los valores de esos elementos, de forma consistente. Iannella opina que la interoperabilidad sintáctica no es un reto técnico pero requiere un nivel de consenso y estandarización.
- *Estructura*: La estructura describe los mecanismos para agrupar elementos de metadatos, así como las reglas sobre cómo deben codificarse los valores para cada elemento. Es similar a la sintaxis, pero conlleva la infraestructura para que metadatos heterogéneos soporten diferente semántica y sintaxis.
- En cuarto lugar, este autor señala que el problema de la interoperabilidad es la *práctica*, entendiendo por práctica las pautas en la aplicación real de metadatos (lo que hemos denominado en otras ocasiones masa crítica de un formato), que se

¹²⁴ TEI Header: <http://www.tei-c.org/P4X/HD.html>

¹²⁵ Sin embargo, como comentamos al hablar de RDF, son tipos de propiedades (*property types*) diferentes. Cfr. 5.2.3.

puede concretar en recomendaciones de uso, e incluso, como hemos visto, en normas¹²⁶.

- A estos niveles que señala Iannella nosotros, en el ámbito de las bibliotecas digitales, añadiríamos además la interoperabilidad entre vocabularios o la interoperabilidad entre metadatos dependientes del contenido.

El nivel de interoperabilidad real de un modelo concreto de metadatos se determinará por la medida en que ese modelo soporte estos cuatro aspectos señalados por Iannella pero, sobre todo, es indispensable la interacción semántica para una efectiva recuperación entre formatos. No obstante, hablar de interoperabilidad entre sistemas de metadatos o de la construcción de *crosswalks* para la interacción entre distintos modelos, tiene muchas más implicaciones para la recuperación de información de las que podemos tratar aquí ahora debido a la propia amplitud y heterogeneidad del espacio de información electrónica en la Web y la inadecuación manifiesta de su protocolo base HTTP (*HyperText Transfer Protocol*) para acceder a toda la amplitud del concepto de bibliotecas digitales tal y como lo hemos definido en el apartado anterior (8.2.).

Cada biblioteca digital utiliza su propio modelo de metadatos adecuado al tipo de información que alberga la biblioteca, e incluso puede optar por distintos esquemas para cada tipo de DLO o para cada colección de objetos electrónicos que mantenga. La complejidad y variedad de los distintos modelos de metadatos dificulta el intercambio de información, a pesar de que se trate de información ampliamente estructurada en XML/SGML. Esta dificultad se magnifica si pensamos en las bibliotecas híbridas y en la necesidad de integración de las colecciones digitales organizadas según un esquema (*schema*) de metadatos particular, con las colecciones tradicionales codificadas en MARC, en un mismo contexto institucional que pretende

¹²⁶ Vid. Capítulo 6.

dar un servicio integral de interfaz Web; y más aún si proyectamos la búsqueda de información entre distintas bibliotecas digitales en la WWW.

De forma sistemática se plantean dos entornos generales (E_1 y E_2) de recuperación de información en la Web que precisan interoperabilidad entre los formatos en que se fundamentan:

- 1) Por un lado, el entorno de la información marcada (E_1), es decir, información en HTML/XML en la que los metadatos, independientemente del formato que sigan, se encuentran embebidos en la propia sintaxis del documento (a través de la etiqueta <META>) y almacenados, de forma conjunta, con el objeto que describen. A este contexto pertenecerían los metadatos que hemos tratado con mayor profundidad en esta investigación, los que hemos descrito en el apartado conceptual como metadatos restringidos al contexto de la información electrónica distribuida en texto completo a través de la Web¹²⁷.
- 2) Y por otra parte, el entorno de la información no-marcada¹²⁸ (E_2), es decir, toda aquella información a la que se puede acceder a través de la Web pero que no está en HTML. Dentro de E_2 , podemos diferenciar además dos subentornos:
 - a) Y el entorno de información formado por DLOs no-textuales o no susceptibles de comportar un marcado descriptivo almacenado con el propio DLO. Nos referimos por ejemplo a los documentos PDF, y a todos los formatos de información visual o multimedia.
 - b) El entorno de la información de las bases datos bibliográficas, como por ejemplo, WebPACs basados en registros MARC, esto es, información que

¹²⁷ Cfr. 3.2.

¹²⁸ En el taller Metadata Summit organizado por el RLG en 1997, donde se trató la posibilidad de ampliar la utilidad del DC a información no basada en HTML, se denominó a este mismo entorno *Información no-HTML*. Willy Cromwell-Kessler and Ricky Erway. *Op. cit.*, <http://www.rlg.org/meta9707.html> Sin embargo, aquí utilizamos la expresión más general (información no-marcada) que engloba más realidades y que implica una actualización frente a la realidad y proyección estructural de la WWW.

no está estrictamente en la Web, pero sí sus datos representativos (la información secundaria en sentido tradicional). En cualquier caso está información está basada también en metadatos pero en el sentido del concepto de "todo metadatos", que interpretaban cualquier descripción de la información como metadatos.

Según esto, desde nuestro punto de vista, la interoperabilidad semántica para la recuperación en el entorno de información E_1 y también entre E_1 y E_2 a) puede estar mediatizada por la propia naturaleza integradora de RDF, ya que en ambos entornos la información está constituida por DLOs. De esta forma se pueden crear sistemas de recuperación de información que se fundamenten en la metainformación codificada en RDF, que integre la semántica particular de los distintos modelos de metadatos destinados a describir las particularidades de cada tipo de DLO. Los entornos informativos E_1 y E_2 a), pueden confluír en distintas realidades de bibliotecas digitales o en un acceso globalizado a varias de ellas.

Los distintos niveles de interoperabilidad señalados por Iannella pueden traducirse en documentación accesible para el hombre (recomendaciones, manuales de sintaxis, especificaciones de formatos, etc.), pero en entornos de información exclusivamente digital, esos niveles deben ser también legibles y entendibles por ordenador. Por ello, y según las expectativas puestas en el RDF, es el único modelo —más bien, metamodelo o metaformato—, que permite representar de una forma entendible por el ordenador el esquema de metadatos que detalla los elementos que pertenecen a un formato determinado, que es fundamental, según varios autores¹²⁹,

¹²⁹ V.gr. Stuart Weibel and Juha Hakala. DC-5. The Helsinki Metadata *Op. cit.*, <http://www.dlib.org/dlib/february98/02weibel.html>; Renato Iannella. Mostly Metadata... *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA1998>; Eric Miller, An introduction to Resource... *Op. cit.* <http://www.dlib.org/dlib/may98/miller/05miller.html>. Desde sus orígenes (1998), todos los autores que hablan de RDF, se refieren a esta facilidad implícita para representar esquemas de distintos formatos.

para la interoperabilidad entre ellos. RDF es un conjunto de reglas para crear semántica, y el esquema RDF (RDFS) una forma de crear vocabularios.

Otra dimensión del problema de la interoperabilidad entre metadatos, se produce cuando se trata de recuperar información entre formatos de metadatos de E₁ y E₂b). Es decir, al tratar de globalizar las búsquedas dirigidas tanto a información marcada a través de XML o HTML, como a información no-marcada, aunque sí estructurada, por ejemplo, utilizar un OPAC como pasarela de búsqueda de información de la Red o, a la inversa, utilizar un sistema de búsqueda basado en metadatos asociados a DLOs, a bases de datos bibliográficas codificadas en MARC. Esto se conseguirá a través de la creación de *crosswalks* o mapeo entre los formatos de codificación de metainformación implicados, es decir, a través de *una correspondencia programada entre información en diferentes formatos, usada para la traducción y la conversión de datos*¹³⁰.

Esta necesidad de integrar las búsquedas en catálogos bibliográficos constituidos en MARC y diferentes pasarelas temáticas e incluso sistemas de recuperación de información general en la WWW, es improporrogable en el caso de las bibliotecas híbridas. Como ya dijimos, algunos autores hablan de la biblioteca híbrida como una etapa previa a "la" biblioteca digital¹³¹, cuya consecución dependerá más de que se opere un cambio cultural que del propio desarrollo tecnológico. Además de las bibliotecas híbridas constituidas por una decisión formal de formar un servicio de información específico adaptado a los usuarios de una comunidad que sirva de puente entre la biblioteca convencional y la digital, ya hemos mencionado, que existe una tendencia generalizada en todas las bibliotecas a establecer una política de aproximación u organización que enfrente la información Web. Esta circunstancia

¹³⁰ Cfr. Glosario, *crosswalks/crossroads*.

¹³¹ Charles Oppenheim and Smithson, Daniel. What is the Hybrid Library? *Journal of Information Science*, 1999, vol. 25, nº 2, 97.

hace que algunos autores, como Michael Gorman¹³², señalen los materiales de los que debe constar la colección de una biblioteca moderna:

- *Objetos tangibles (libros, grabaciones, mapas, vídeos, etc.)* propiedad de la biblioteca y resguardados en un lugar físico específico.
- *Recursos (electrónicos) tangibles*¹³³ (*CD-ROMs, etc.*) existentes en la biblioteca y disponibles, normalmente, para su uso dentro de ésta.
- *Objetos tangibles, (libros, etc.)* propiedad de otras bibliotecas pero accesibles a través del préstamo interbibliotecario y de proyectos de intercambio bibliográfico como bases de datos colectivas.
- *Materiales remotos intangibles (electrónicos)* que no existen físicamente en una biblioteca en cuestión, pero sin embargo son accesibles, a través de Internet.

El mayor reto de la interoperabilidad es el entorno híbrido donde confluye el acceso a todos los entornos informativos E₁-E_{2a}) y E_{2b}). Una de las tareas más difíciles es reconciliar el hecho de que la biblioteca tendrá que gestionar sistemas de recuperación de información local mientras crea y mantiene interfaces de acceso global, o una recuperación híbrida entre lo tradicional y lo digital. Hasta ahora, en las bibliotecas modernas¹³⁴, el problema de conseguir la interoperabilidad semántica entre fuentes de datos heterogéneas se había conseguido a través de Z39.50. Sin embargo, la utilidad de Z39.50 para alcanzar la interoperabilidad entre metadatos ha dado lugar a opiniones muy confrontadas que nos permiten señalar dos tendencias e

¹³² Michael Gorman. ¿Metadatos o catalogación?: un cuestionamiento erróneo. En: *Internet, metadatos y acceso... Op. cit.*, p. 2

¹³³ *Ibid.* A pesar de que el traductor de este documento, Filiberto F. Martínez, traduce *recursos (electrónicos) intangibles* [sic.], resulta obvio (a pesar de que no lo hemos podido confrontar con el texto original, pues no han aparecido publicadas las ponencias originales en inglés) que el autor, Gorman, se refiere a los recursos electrónicos en soporte *tangible*.

¹³⁴ Entendiendo por biblioteca moderna-tipo, aquella conectada a Internet y que tiene su WebPAC y que aprovecha la comunicación en Red para crear consorcios, y entidades de colaboración, normalmente u originalmente basadas en la compatibilidad MARC.

incluso etapas sobre la aplicación de Z39.50 para el acceso global entre sistemas de información basados en distintos modelos de metainformación, incluyendo entre ellos el MARC y la concepción de bibliotecas híbridas.

Por un lado distinguimos una tendencia o etapa que rechaza Z39.50 para dicho fin. Así, por ejemplo, Ward y Wood¹³⁵ mostraban su descrédito en este protocolo de intercambio de información en estructuras de metadatos, debido a su substrato bibliotecario, argumentando además que se desarrolló principalmente para facilitar el acceso a bases de datos de referencias, más que para acceder a los recursos en sí mismos. Por su parte, Amanda Xu¹³⁶ señala las limitaciones de este sistema de búsqueda y recuperación, denunciando su poca flexibilidad para manejar todos los tipos de metadatos, amén de los problemas de lentitud en las respuestas, y el hecho de que se base en sesiones de consulta. Catherine Leloup¹³⁷ denuncia también este último aspecto y además censura el lenguaje de consulta de Z39.50, considerándolo algo retrasado con respecto a otros productos del mercado y, la falta de consenso en la especificación de juegos de caracteres.

Por otra parte, podemos señalar otra tendencia o etapa de Z39.50 que coincide también con la segunda fase en el desarrollo de los metadatos iniciada, en nuestra opinión en 1998. Esta tendencia desarrollada, sobre todo en los dos últimos años (2000-2001), aboga por una redefinición de este estándar ISO 233950-1998 y destaca sus fortalezas para la recuperación de información en entornos heterogéneos, ampliando su aptitud al mundo de los metadatos asociados a DLOs. Así, se ha

¹³⁵ Nigel Ward and Andrew Wood. *Emerging Technologies for Networked Information Discovery: Beyond Z39.50 Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA98>

¹³⁶ Amanda Xu. *Metadata Conversion and the Library OPAC* [documento HTML]. Massachusetts: Institute of Technology, rev. 30 de enero de 1998; The Haworth Press, 1997. Disponible en: <http://web.mit.edu/waynej/www/xu.html> (consultado el 24 de julio de 1998).

¹³⁷ Catherine Leloup. *Op. cit.*, p. 120-121.

constituido ZING¹³⁸ (*Z39.50 International Next Generation*), una iniciativa trata de adaptar todo el capital intelectual de Z39.50, desarrollado en los veinte últimos años, y además recoge todas los intentos (v. gr. ZOOM, SRW, etc.) que permitan adaptar dicho protocolo al nuevo entorno Web-XML, de tal forma que sea más atractivo para los proveedores, vendedores y usuarios de información. En esta fase encontramos opiniones de Z39.50, incluso provocadoras como las de Taylor¹³⁹, que afirma que iniciativas como el DC o RDF, implican la abstracción de aspectos consolidados en el mundo de Z39.50.

A pesar de que el protocolo Z39.50 está mejorando sus capacidades y continuará adaptándose a las necesidades de las bibliotecas virtuales, no reúne todos los requisitos para poder considerarlo un "metabuscaador de bibliotecas digitales" que recupere información de calidad de entre todos los depósitos documentales del la Red. Por ello, se crean nuevos *crosswalks* entre distintos sistemas de metadatos para posibilitar la interoperabilidad entre la metainformación generada para el entorno 1 y en el entorno 2. Este es el caso de la traducción de cualquier modelo de metadatos al formato USMARC¹⁴⁰ a través de un sistema de conversión¹⁴¹ que analice y *mapee* los elementos comunes entre los dos esquemas (p. ej. 100=Creator). Esta forma de funcionamiento es la que se advierte por ejemplo en CORC que, a través de la tecnología desarrollada en el proyecto MANTIS y fundada en XML/RDF, permite

¹³⁸ Z39.50 International Next Generation <<http://www.loc.gov/z3950/agency/zing/zing.html>> Las ideas de este grupo fueron presentadas en el último congreso del ZIG celebrado en octubre de 2001, por Ray Denenberg: <http://lcweb.loc.gov/z3950/agency/zng/zng.ppt>

¹³⁹ Mike Taylor. *Op. cit.*, <http://www.miketaylor.org.uk/tech/esb.html> Vid. Capítulo 5, nota 216.

¹⁴⁰ El hecho de convertir los metadatos al formato USMARC y no a otros MARC-nacionales, se debe a que es el único que ha integrado el campo *de localización y acceso electrónico* (856) que, como explicamos en el apartado 4.1. de este trabajo, permite conectar directamente los registros USMARC con los recursos electrónicos que describe gracias a una nueva generación de OPACs integrados en la Web (WebPACs).

¹⁴¹ Por ejemplo, como ya hemos mencionado al hablar de las herramientas y aplicaciones vinculadas a los metadatos (4.3.2.4.), el Nordic Metadata Project ha desarrollado d2m (*Dublin Core to Marc Converter*) para convertir DC a formato MARC. Vid. Figs. 51 y 52.

trabajar tanto en MARC como en el formato DC, aunque siempre con información perteneciente al entorno E₁.

A pesar de que esta última posibilidad de *crosswalks* es factible, defendida por algunos autores como Xu¹⁴², e incluso más cómoda para sectores bibliotecarios ampliamente vinculados y familiarizados con el formato MARC, existen críticas sobre la posibilidad de combinar provisionalmente los dos paradigmas de información¹⁴³. Queremos resaltar además, lo que ya apuntamos al hablar del SGML y XML como formatos de intercambio bibliotecario, MARC es un formato menoscabado en el entorno de la información Web, constituido para el universo Gutenberg de información o el mundo de información tangible que, si bien puede llegar a describir el nivel de procedimiento de los documentos electrónicos, dista mucho de ser un formato apto para definir su estructura conceptual y avalar la recuperación de información relevante en Internet¹⁴⁴.

En cualquier caso, desde una perspectiva bibliotecaria, es preciso conseguir la interoperabilidad, tanto entre sistemas basados en modelos exclusivamente diseñados para la información estructurada de la Red¹⁴⁵ y entre éstos y sistemas de información

¹⁴² Amanda Xu. *Op. cit.*, <http://web.mit.edu/waynej/www/xu.html>

¹⁴³ Por ejemplo Gradmann expresó su escepticismo en el congreso de la IFLA de 1998, con estas palabras: ...como en la propuesta de Xu de utilizar el OPAC de la biblioteca como pasarela de acceso a un archivo de metadatos ...*tengo mis dudas sobre su aplicación práctica inmediata*... Stefan Gradmann. Cataloguing vs. Metadata: Old Wine in New Bottles? [documento HTML]. En: *IFLA General Conference (64. 1998. Amsterdam)*. París: IFLA; INIST, rev. 23 de junio de 1998. Disponible en: <http://ifla.inist.fr/ifla/IV/ifla64/007-126e.htm> (consultado el 6 de julio de 1998). No obstante este autor también reconoce que combinar la automatización bibliotecaria y las técnicas de información en Internet, es una línea de investigación interesante para los profesionales de la información. Deacon, más categórico, afirma que *los estándares de metadatos deben concentrarse en la recuperación de información en Internet (resource discovery) y no intentar interoperar con MARC*. Prue Deacon. *Op. cit.*, p.42.

¹⁴⁴ *Cfr.* 4.1. y 6.1.

¹⁴⁵ Por ejemplo Renardus <<http://www.renardus.org>>, es un proyecto de colaboración financiado en el marco del IST del 5PM, que aspira a la interoperabilidad entre distintas bibliotecas virtuales temáticas (*subject gateways*) europeas. Actualmente (2001) se está trabajando en el análisis de

[cont.]

basados en el formato MARC¹⁴⁶, que se producirá a través de alguna o varias de las siguientes opciones:

- a) Formatos más ricos semánticamente como XML.
- b) Modelos aptos para reflejar el esquema de metainformación como RDF¹⁴⁷.
- c) Soluciones de interoperabilidad basadas en la arquitectura (v. gr. CORBA-*Common Object Request Broker Architecture*¹⁴⁸).
- d) A través de la disseminación del contenido de los metadatos, con ayuda de protocolos nuevos como OAI (*Open Archives Initiative*).

El problema de integrar, controlar e intercambiar semántica asociada a los DLOs de las diversas bibliotecas digitales, esto es metadatos, se está convirtiendo en algo cada vez más importante e ineludible. Las iniciativas que están surgiendo en este sentido así lo demuestran. Queremos resaltar como ejemplo la OAI, que tiene el objetivo de garantizar la interoperabilidad entre bibliotecas digitales o archivos de información editada electrónicamente. La OAI ha establecido un conjunto de especificaciones relativas a la interoperabilidad para facilitar el desarrollo de

modelos para compartir metadatos, de acuerdo con soluciones técnicas y otros estándares y en el desarrollo piloto de un sistema para el acceso global e interdisciplinar entre esas bibliotecas digitales europeas. Para más información sobre este proyecto, Vid. Lesly Huxley. Renardus: Fostering Collaboration Between Academic Subject Gateways in Europe. *Online Information Review*, 2001, vol. 25, nº 2, p. 121-127.

¹⁴⁶ En este tipo de interoperabilidad en contextos de información híbridos, destaca el proyecto MALIBU (*MANaging the hybrid Library for the Benefit of Users*) <<http://www.kcl.ac.uk/humanities/cch/malibu>> que ha desarrollado un prototipo de *search engine* para buscar simultáneamente entre distintos tipos de información *online*, tanto OPACs como revistas electrónicas, sitios Web o archivos digitales.

¹⁴⁷ Como explicamos al hablar de RDF, en la concepción de este metamodelo de metadatos subyace una idea similar a la del formato GRS-1 del Z39.50. Cfr. 5.2.3.

¹⁴⁸ Nigel Ward and Andrew Wood. *Op. cit.*, <http://www.dstc.edu.au/RDU/reports/VALA98>

servicios creados por terceros. Entre estas sencillas, aunque potentes soluciones, destaca el propio protocolo OAI¹⁴⁹ para la indización y recopilación de metadatos¹⁵⁰ que, desarrollado desde enero de 2001, trata de proporcionar y promover un marco de interoperabilidad independiente de las aplicaciones, que puedan utilizar distintas comunidades de usuarios comprometidos con la publicación de contenidos en la Web. Esto incluye tanto a proveedores de datos, como de servicios, y por tanto, al nuevo subsistema editorial que surge las bibliotecas digitales (*Vid.* 8.2.1)..

No se trata de reemplazar los sistemas existentes o de ignorar un siglo de tradición en la descripción, pero sí de continuar en la construcción de nuevos sistemas metainformacionales, aumentar la comprensión en la creación, mantenimiento y escalabilidad de metadatos a través de esfuerzos cooperativos (como Renardus o IMesh), y aunar la experiencia de la catalogación y de las nuevas técnicas informáticas en la creación de estándares y metadatos de cara a un futuro informativo imperiosamente digital. En cualquier caso, teniendo en cuenta que los sistemas de metadatos difieren tanto en su amplitud, como en su cobertura y finalidad, el establecimiento de *crosswalks* y técnicas de mapeo entre distintos modelos resulta complejo porque, al igual que ocurre en una traducción lingüística, nunca es exacta la traducción entre metadatos. Sin embargo, partiendo de que la consecución de interoperabilidad entre sistemas, está abocada a facilitar la recuperación de información en Internet a través de la gran variedad de recursos accesibles en las distintas bibliotecas digitales "locales", es importante que las decisiones se tomen de manera uniforme, confiando a los metadatos, más que una

¹⁴⁹ Para una información detallada sobre la última versión del protocolo OAI, *Vid.* *The Open Archives Initiative Protocol for Metadata Harvesting. Protocol Version 1.1 of 2001-07-02* [documento HTML]. Herbert Van de Sompel, Carl Lagonze, eds. OAI, 20 de junio de 2001. Disponible en: http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html (consultado el 1 de diciembre de 2001).

¹⁵⁰ *Cfr.* Glosario, *harvesting/gathering*.

mera funcionalidad descriptiva, una organización inteligente de los sistemas de información distribuidos y, por ende, una recuperación eficaz.

Para una verdadera interoperabilidad tanto dentro de una misma institución como fuera de ella, se requiere un conjunto de reglas y estándares universalmente aceptados así como una infraestructura semántica para esas normas análoga a las ISBD, o a un vocabulario controlado universal. Esta tarea de normalización sería mucho más ambiciosa que la que enfrentaron los creadores de las normas catalográficas bibliotecarias, sin embargo la extensibilidad de XML, incluyendo la facilidad para establecer enlaces, implica que se puede inventar un sistema de información estructurada que permita el intercambio de información. Por otra parte, para que se practique la tendencia hacia la interoperabilidad en bibliotecas digitales es necesario además un espíritu de colaboración y cooperación semejante e incluso mayor al que se ha manifestado en las bibliotecas tradicionales de los últimos diez años.

Debemos además hacer referencia a los problemas que se suscitan en torno a la interoperabilidad en el vocabulario. Para potenciar el deseo de recuperación global en las bibliotecas digitales es preciso contar no sólo con modelos de metadatos que interoperen en un nivel técnico, sino que el contenido de esas estructuras de información sea lo más uniforme (quizás, normalizada) posible. Para ello, no sólo el desarrollo tecnológico será suficiente para mejorar la recuperación de información, sino también una representación ortodoxa del conocimiento distribuido, lo que, hasta hoy, sólo es posible mediante el lenguaje. Según esto, y ante la imposibilidad de desarrollar un mega-tesauro válido para todas las bibliotecas virtuales, es preciso conseguir también interoperabilidad de vocabularios, *conectar tesauros en la red de forma que se pueda acceder a un tesauro en un sitio y usarlo para indizar o buscar recursos en otro sitio, lo que significaría un desarrollo de estándares en términos de*

*semántica de relaciones tesaurales, de igual forma que la sintaxis de consulta de tesauros*¹⁵¹.

Entre las visiones locales y globales, que hemos manejado a lo largo de todo este capítulo, ese mapeo puede obtenerse a través de la construcción de un "Tesoro Común" de relaciones intra/inter-esquemas (*schemes*), capaz de reconciliar diferentes representaciones de conceptos similares. Este es el caso de proyectos como MOMIS¹⁵², o HILT (*High Level Thesaurus*¹⁵³), este último, financiado por el Programa de Soporte a la Investigación en Bibliotecas (RSLP) y el JISC británico, trata de proponer recomendaciones o *best practices*¹⁵⁴ relacionadas con el uso de sistemas de organización del conocimiento (tesauros y otras terminologías), tanto los orientados a su procesamiento por máquina (v. gr. XTM, Voc-ML o RDFS) como a interfaces tradicionales (como el AAT), con la finalidad de conseguir la interoperabilidad al nivel del vocabulario, para un acceso por materias integral entre distintas bases de datos y servicios de información electrónica distribuida (bibliotecas, archivos y museos digitales)¹⁵⁵.

¹⁵¹ Ron Davies se expresaba en estos términos en 1998, durante la presentación del taller sobre aplicación de herramientas terminológicas y de clasificación para el desarrollo de colecciones digitales y la búsqueda basada en la Red. *Application of Terminology and Classification Tools for Digital Collection Development and Network-based Search: ACM DL Workshop (1998. Pittsburgh)* [documento HTML]. Alexandria Digital Library. Alexandria Digital Library. Santa Bárbara: University of California, rev. 13 de julio de 1998. Disponible en: http://www.alexandria.ucsb.edu/~lhill/nkos/dl98_workshop_preliminary.html (consultado el 28 de agosto de 1998).

¹⁵² MOMIS (*Mediating system environment fOr Multiple Information Sources*): <http://sparc20.ing.unimo.it/Momis>

¹⁵³ Susannah Wake, Dennis Nicholson. *Op. cit.*, Susannah Wake, Dennis Nicholson. HILT-High Level Thesaurus Project [documento HTML]. *D-Lib Magazine*, September 2001, vol. 7, n° 9. Disponible en: <http://www.dlib.org/dlib/september01/wake/09wake.html> (consultado el 15 de octubre de 2001).

¹⁵⁴ *Vid.* Glosario.

¹⁵⁵ Un informe completo de los estudios realizados en el seno de HILT es: Rosemary Russell and Michael Day. *HILT: High-Level Thesaurus project: Automated and Manual Approaches to the Provision of Thesauri and Subject Vocabularies* [documento HTML]. Bath: University, UKOLN, 5 de octubre de 2001. Disponible en: <http://www.ukoln.ac.uk/metadata/hilt/interfaces> (consultado el 15

[cont.]

La WWW es un espacio de comunicación sin precedentes que permite que las máquinas se comuniquen entre sí gracias a aspectos de la evolución tecnológica como los lenguajes de marcado y la asunción rápida de estándares que hacen evolucionar la Web. Esa comunicación entre máquinas potencia la globalización de la información y la comunicación entre personas, sin embargo, queda por resolver la comunicabilidad entre las máquinas y los humanos, de tal forma que los ordenadores puedan aprehender el sentido de la lengua humana, esto sólo se conseguirá a través de una Web más semántica y de la interoperabilidad entre sistemas. Técnicamente está todo en su sitio para implementar ese modelo de estructuración de la información que permita exportar a la Web las reglas de cualquier sistema de representación del conocimiento. XML está bien definido, RDF proporciona un marco uniforme para dotar de significado a la estructura de los documentos de tal forma que lo procesen las máquinas, los software actuales permiten facilidades de búsqueda a una velocidad y complejidad que iguala los mejores sistemas tradicionales de las bases de datos y, los sistemas de clasificación, las ontologías y la organización del conocimiento para la WWW también están a punto. Sólo resta esperar al siguiente estadio en la madurez en la organización y la estructuración de la información para convertir la Web en la Web Semántica.

De igual forma, a nivel político existe una tendencia uniforme para promover el acceso universal a la información. Así, por ejemplo, en el ámbito europeo, la Comisión y su Dirección General para la Sociedad de la Información, han instituido los principios para lograr la interoperabilidad y el acceso global a las bibliotecas digitales, a través de lo que ya se conoce como *Los principios de Lund*¹⁵⁶, instando a

de octubre de 2001). Recordemos además que esta visión de la recuperación temática basada en metadatos orientados a la materia la perfilamos en el apartado 6.4. No obstante HILT aspira a sentar las bases de una recuperación por materias inter-dominios, y por ello inter-terminológica. Cfr. Capítulo 6, nota 122.

¹⁵⁶ Vid. *Report of a Formal Meeting held at Rauol Wallenberg Centre Lund, Sweden. 4 April 2001* [documento PDF]. Cordis, Digicult, abril de 2001. Disponible en: <ftp://ftp.cordis.lu/pub/ist/docs/digicult/lund-report.pdf> (consultado el 23 de mayo de 2001).

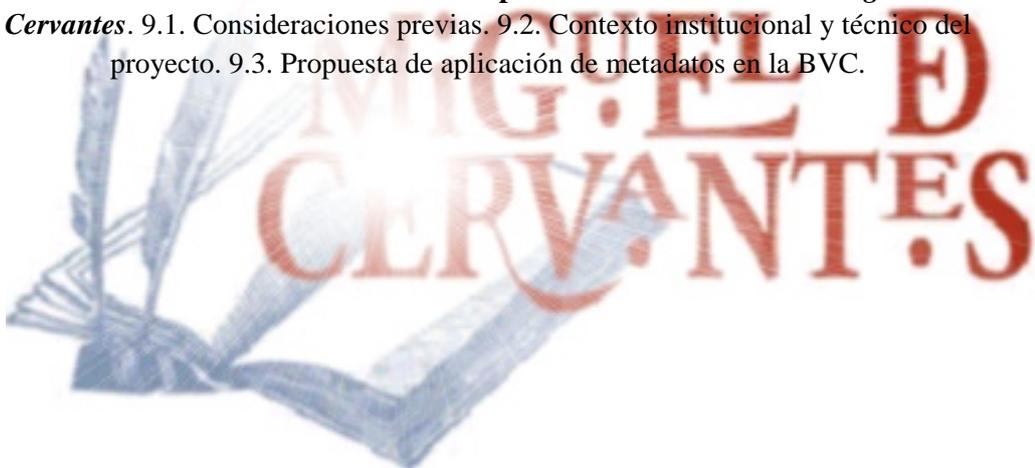
mejorar la calidad y usabilidad de los contenidos, promover el acceso unificado de los ciudadanos y sensibilizar sobre los problemas de conservación, a través de la elaboración de acuerdos sobre normas de interoperabilidad y de directrices para la preservación digital y longevidad de los contenidos. Igualmente se reconoce la necesidad de favorecer la interoperabilidad y la recuperación continuada de recursos a través de trabajos sobre metadatos, registros y esquemas, así como lograr una coherencia terminológica que favorezca el multilingüismo y la preservación de la identidad cultural.

PARTE IV:

Proyecto de aplicación de metadatos en bibliotecas digitales: Gestión de una colección digital a través de un sistema descriptivo basado en metadatos

BIBLIOTECA VIRTUAL

9. Diseño de un sistema de metadatos para la Biblioteca Virtual *Miguel de Cervantes*. 9.1. Consideraciones previas. 9.2. Contexto institucional y técnico del proyecto. 9.3. Propuesta de aplicación de metadatos en la BVC.



CAPÍTULO 9:

DISEÑO DE UN SISTEMA DE METADATOS PARA LA BIBLIOTECA VIRTUAL MIGUEL DE CERVANTES

Quienes lo imaginan sin límites, olvidan que los tiene el número posible de libros. Yo me atrevo a insinuar esta solución del antiguo problema: La biblioteca es ilimitada y periódica. Si un eterno viajero la atravesara en cualquier dirección, comprobaría al cabo de los siglos que los mismos volúmenes se repiten en el mismo desorden (que, repetido, sería un orden: el Orden). Mi soledad se alegra con esa elegante esperanza.

Jorge Luis Borges. *La biblioteca de Babel*, 1941.

Como anticipamos, la última parte y el último capítulo de esta tesis doctoral tratará de absorber todas las reflexiones y conclusiones a las que hemos llegado en esta investigación a través de una propuesta de aplicación de metadatos en la Biblioteca Virtual *Miguel de Cervantes* (BVC) <<http://www.cervantesvirtual.com>>, la mayor biblioteca digital en lengua hispana en Internet, que todavía no utiliza esta tecnología como fundamento para la recuperación de información en su colección digital.

Este capítulo estará constituido pues, por el diseño de un proyecto basado en el análisis de los propios objetivos de esta biblioteca, sus flujos de trabajo, el personal con el que cuenta, la tecnología en la que se fundamenta, etc. y en definitiva, en su realidad como servicio de información digital. Se trata de hacer un planteamiento analítico para la aplicación de metadatos en un supuesto real, en el que haremos confluir, ahora de una forma objetiva y aplicada, todas las ideas relacionadas, tanto con concepto de metadatos como con su aplicación en bibliotecas digitales y a toda la problemática que hemos tratado en esta tesis doctoral, sin perder de vista los

parámetros que implica la nueva Sociedad de la Información: universalidad informativa y acceso global. Por todo ello, tratamos de conjugar en el proyecto tanto las técnicas bibliotecarias de organización del conocimiento, con las nuevas tecnologías de estructuración de la información.

9.1. Consideraciones previas

Como complemento práctico de aplicación de los metadatos en el contexto de las bibliotecas digitales, en consonancia con la tesis defendida aquí —la operatividad de los metadatos en conjuntos abarcables de información electrónica o bibliotecas digitales, donde la metainformación se entiende como el fundamento de la organización de dichas bibliotecas— podríamos haber planteado al menos tres tipos de trabajos:

1. La implementación de metadatos en una biblioteca digital que fuese al mismo tiempo la editora de su colección digital y por ende, pudiera plantearse el proceso de asignación de metadatos como parte del flujo de trabajo de la producción y organización de documentos digitalizados.
2. Un supuesto de creación de una *subject gateway* en la que aplicásemos los criterios de selección y organización de una "colección virtual", creando una base de datos y un sistema de recuperación de información basados en la teoría creación de metadatos por agentes externos (*retro-codificación*¹ o asignación de metadatos por gestores externos al creador del recurso²).
3. La creación de una biblioteca híbrida donde planteásemos la integración de una colección digital y una colección tradicional, poniendo a prueba la

¹ Jennie Thornely. *Op. cit.*, p. 33. *Vid.*4.2.

² *Third party creators* en la denominación de Dempsey y Heery. Lorcan Dempsey and Rachel Heery. *Metadata: A Current View...* *Op. cit.*, p. 165. *Cfr.* 4.2.

interoperabilidad de un modelo de metadatos con el catálogo tradicional realizado en MARC.

De estos planteamientos, hemos tomado el primero como marco de aplicación de esta investigación. Entre otras múltiples circunstancias que comentaremos a lo largo de este capítulo, la Biblioteca Virtual *Miguel de Cervantes* es una biblioteca digital en sentido estricto ya que digitaliza y crea su propia colección, a la par que la organiza y da acceso a ella. A pesar de que la creación de una pasarela temática (2) o de una biblioteca híbrida (3) *ex novo* nos otorgaría más autonomía en el diseño de un modelo de recuperación de información al crear los metadatos *a posteriori* (según la terminología que manejamos en este trabajo, subrogados o representaciones post-publicación del DLO³), conllevaría otros problemas o conflictos con los objetivos de esta tesis. En el caso 2, la creación de una *subject gateway*, tendríamos que aplicar además principios de selección y evaluación de recursos Web, aspecto que excluimos *ex professo* de esta investigación. Y en el caso 3, la formación de una biblioteca híbrida, además de la selección, el planteamiento estaría demasiado mediatizado a una herramienta de software (SIGB) y a un catálogo preexistente.

Sin embargo el primer planteamiento nos permite acercarnos más a muchas de las ideas defendidas aquí con relación a la metainformación y al concepto más estricto de biblioteca digital, ya que la Biblioteca Virtual *Miguel de Cervantes* se fundamenta la creación de su colección en la digitalización, marcado, organización y publicación de los documentos. Se puede aplicar por completo la concepción de la biblioteca como concepto y como sistema en el contexto digital de la comunicación global, existiendo además el nuevo subsistema que hemos denominado "editorial" y que permite plantear una visión de los metadatos como codificación ECIP (*Electronic Cataloging in Publication*) que es la visión más aceptada y defendida.

³ Según el concepto de Qin *post-publishing representation*. Jian Qin. *Op. cit.* p. 84; y según el concepto de *subrogate* o sustituto documental que tratamos en el capítulo 3. *Cfr.* Capítulo 3, nota 107, Glosario, *surrogate*, *subrogate*.

En la BVC, confluyen además, en mayor o menor medida, todos los objetivos señalados por Schwartz⁴ que, desde un punto bibliotecario, debe tener una biblioteca digital y que discutimos a continuación incidiendo en aquellos aspectos que mejorarían según nuestra visión de los metadatos para la recuperación de información:

- Aunque no sirve a una comunidad definida de usuarios, tiene la finalidad, implícita en la concepción de su filosofía, de servir a una comunidad universal. Esa comunidad universal puede seccionarse en distintos colectivos de interés como filólogos, literatos, historiadores, personal académico vinculado a las Humanidades, etc. La utilización de metadatos normalizados (RDF) permitirá tanto la división de la colección según comunidades de usuarios como la ampliación del acceso a través de la generación de *crosswalks* con otras bibliotecas digitales.
- No es una entidad aislada porque el proyecto está vinculado a la Universidad de Alicante y cuenta con el refrendo de una comunidad científica internacional, además tiene como mecenas al Banco de Santander. Todo ello permite vincularla al conjunto de objetivos de las distintas organizaciones implicadas en su gestión.
- Tiene una estructura organizacional perfectamente definida, un organigrama y una correcta definición de los flujos de trabajo, que hace viable la inserción del planteamiento de la metainformación.
- La BVC implica el acceso, el aprendizaje e incluso la investigación, que nos permiten vincularla a la nueva concepción de la biblioteca como sistema, dentro de la comunicación social global y de la educación, que proponemos en esta tesis (*Vid.* Fig. 60). Acerca de esta forma, a cualquier usuario de la Red, la posibilidad de acceso al patrimonio cultural hispanoamericano y al aprendizaje a partir de él.

⁴ Candy Schwartz. *Op. cit.*, <http://web.simmons.edu/~schwartz/530-defs.html> Cfr. 8.2. donde enunciamos las características de las bibliotecas digitales en abstracto, recogidas por Schwartz.

Además incluye una serie de servicios y una dinamización de la biblioteca que la comprometen aún más con la educación.

- En cuanto al aprovechamiento de sus recursos tecnológicos y humanos, utiliza al máximo ambos, sin embargo la aplicación de metadatos permitirá también explotar los recursos bibliotecarios adaptándolos a su potencial tecnológico.
- El acceso a los contenidos y los documentos digitales que alberga es bastante rápido, e inclusive, en algunos casos, contempla varios modos de acceso⁵.
- El acceso a la biblioteca *Virtual Miguel de Cervantes* es totalmente libre y gratuito. Incluso, los investigadores, previa solicitud o convenio, pueden acceder a las obras que aún no se hayan publicado o estén en proceso de corrección.
- Posee y controla sus recursos a todos los niveles, almacenando, gestionando y controlando su colección digital de forma centralizada.
- En cuanto a la formación de la colección⁶, es amplia, pero previsiblemente finita (30.000 obras, DLOs), además persistente (no hay ningún enlace "roto" ni ningún aspecto no operativo desde su creación), y contiene diferentes formatos, e incluso objetos de información digital que no se pueden conseguir fácilmente por otros medios, como son, la literatura gris y las tesis doctorales. También contiene objetos originalmente digitales si entendemos como DLOs también, toda la información que generan en torno a su actividad (los contenidos, no las obras propiamente dichas, de sus portales temáticos, noticias, boletín *Dulcinea*, etc.)

⁵ Además esta afirmación es cierta si consideramos como un modo de acceso, la adaptación de sus servicios al lector de pantalla JAWS <http://www.freedomscientific.com/fs_news/nr_JAWS401.asp> para invidentes. Aunque desde el punto de vista de la accesibilidad WAI del consorcio, la página que proporciona ese modo de acceso no es "accesible", sí se puede acceder a algunos de los servicios de la BVC a través de la aplicación JAWS. Este acceso se encuentra en: <http://www.cervantesvirtual.com/servicios/mapainv.shtml>

⁶ Para una visión más clara de la colección y de la diversidad de realidades informativas, objetos de información digital, etc. *Vid. infr.*

- En cuanto a la organización de la colección, aspecto en el que centramos nuestra tesis de recuperación de información en bibliotecas digitales, si bien permite recuperar todos sus documentos a partir del catálogo, sus funciones podrían mejorarse, potenciarse y adaptarse a la realidad de las bibliotecas digitales, a través de la aplicación de metadatos.

Una vez elegida esta biblioteca digital (que lo es, en sentido estricto en cuanto posee su colección) a la que vamos a aplicar todas las reflexiones hechas a lo largo de esta investigación, la idea inicial era tomar una muestra de treinta y cinco documentos, seleccionados arbitrariamente de la colección de la BVC⁷ y tomar decisiones de asignación de metadatos y de planificación del sistema de gestión de metainformación en función de ellos. Sin embargo, al ser un proyecto real de implementación de metadatos se prefirió hacer un trabajo total de análisis sobre las características de la BVC teniendo en cuenta todos los factores que confluyen en ella de: volumen, variedad de DLOs y el contexto de producción digital. Para esta decisión influyó la opinión del equipo de la Biblioteca Virtual Cervantes que se planteará, como objetivos del año 2002, la posibilidad de incluir metadatos para la gestión y recuperación de su colección⁸, además de nuestra propia convicción de que la realización de un proyecto sobre una muestra puede ser interesante desde el punto de vista de la investigación *per se*, pero puede traducirse en inviable al trasladar el modelo a la realidad total de un sistema de información al que se puede aplicar potencialmente. Además, la muestra correspondía exclusivamente a documentos textuales y la riqueza y complejidad de la BVC es mucho mayor: posee gran cantidad

⁷ La muestra estaba compuesta por un subconjunto de 35 documentos relativos a tres autores: Dicenta (19 documentos), Valera (12 documentos) y Villoslada (4 documentos) marcados en XML, TEI P4, en función de la DTD reducida de la BVC. No obstante, desechada la posibilidad de crear metadatos para esta muestra, hemos utilizado todos estos documentos para hacer el análisis de los niveles de detalle de la TEIH, así como para comprobar sus códigos fuente en la publicación Web y su recuperación tanto en el sistema local de la BVC, como en el entorno global de la WWW.

⁸ Esta circunstancia es la que nos llevó a hacer una reflexión de carácter general, un trabajo de "analista" sobre toda la BVC desde la perspectiva de biblioteca digital que defendemos en esta tesis más que de implementación efectiva en un conjunto de documentos.

de tipos de DLO, servicios, etc. Por todo ello, lo que haremos aquí será plantear consideraciones generales sobre la aplicabilidad de metainformación en la Biblioteca Virtual *Miguel de Cervantes* y proponer un modelo para integrar la creación de metadatos y la recuperación a partir de ellos, así como una previsión del enriquecimiento de las funciones y de la gestión de la colección digital desde un punto de vista bibliotecario.

Por último es preciso reconocer que, a la hora de diseñar y realizar un proyecto de implementación de metadatos en una biblioteca digital, hay que tener en cuenta que esto supondrá una gran inversión en tiempo, personal y otros recursos bibliotecarios e informáticos, sin embargo, y desde nuestro punto de vista, el beneficio justifica el esfuerzo, máxime cuando la biblioteca digital en la que nos basamos cuenta con una riqueza en infoestructura y una adecuación de infraestructuras informáticas y equipo humano, así como una gran capacidad de investigación e innovación que ameritan, a la par que hacen posible, un tratamiento de la recuperación de información de calidad, adaptado a las últimas tendencias.

9.2. Contexto institucional y técnico del proyecto

A continuación vamos a plantear una aproximación a la BVC, por un lado, desde el punto de vista de la filosofía que infunde el proyecto de biblioteca virtual destacando su importancia en el ámbito de las bibliotecas digitales y en el ámbito de la cultura hispánica, y por otro, desde el punto de vista de la tecnología que utiliza para crear su colección digital, haciendo hincapié sobre todo en la estructuración, a partir del marcado XML, de los documentos, en el nivel de "descripción bibliográfica" y en la recuperación.

Aunque ya hemos hecho alusión a la envergadura del proyecto de la Biblioteca Virtual Miguel de Cervantes, *se trata del mejor proyecto de este tipo existente en el*

*estado español*⁹, aportaremos ahora algunos datos y reflexiones que nos permitan avalar formalmente esta afirmación, así como su elección, en esta tesis, como marco de aplicabilidad de la tecnología de metadatos (entendidos éstos como fundamento estructural y de organización de las bibliotecas digitales, que dotan de "credibilidad" documental a la recuperación de información).

9.2.1. Marco del proyecto: importancia de la BVC (1999-2001)

La Biblioteca Virtual *Miguel de Cervantes* es una iniciativa surgida en la Universidad de Alicante bajo los auspicios del Banco Santander Central Hispano. Ambas instituciones se comprometieron a crear las bases mínimas y plataformas necesarias para lograr que el libro y la lengua española puedan tener una referencia sólida en Internet. El resultado es una biblioteca digital que corrobora ese esfuerzo de poner al servicio de cualquier ciudadano del mundo, la cultura y la lengua hispánicas.

El proyecto comenzó a finales de 1998 con un equipo de cinco personas que realizaron los estudios de ensayo y error para determinar la viabilidad del proyecto y para establecer las herramientas y métodos más adecuados para la producción digital de documentos. Después de esa fase inicial, la producción masiva de "libros digitales", comenzó en marzo de 1999 con un equipo de aproximadamente ochenta personas, hoy más de ciento veinte¹⁰, que trabajarían en la digitalización y corrección

⁹ Esta afirmación tan rotunda no es sólo una opinión nuestra. Diversas publicaciones aparecidas en la prensa nacional e internacional la avalan. En el ámbito profesional, Sorli y Merlo en su reflexión sobre los proyectos de bibliotecas digitales que están constituidos por libros de libre acceso la clasifican también con estas mismas palabras. Ángela Sorli Rojo, José Antonio Merlo Vega. Bibliotecas Digitales (I): Colecciones de libros de acceso público. *Revista Española de Documentación Científica*, 2000, vol. 23, nº 1, p. 99.

¹⁰ En un artículo publicado este año (2001) por Andrés Pedreño, anterior rector de la Universidad de Alicante y principal promotor del proyecto y Alejandro Bia, subdirector de informática y responsable de innovación de la BVC, se destaca que los recursos humanos de la BVC son un grupo de 129 personas entre profesionales y técnicos, de los cuales, un 71% son correctores y marcadores de los textos, un 9% personal encargado de la digitalización, un 10% investigadores y desarrolladores informáticos, un 4% bibliotecarios y el resto (preparadores de textos, administrativos, y responsables del mantenimiento informático). Alejandro Bia, Andrés Pedreño. *Op. cit.*, p. 152.

de textos, así como un equipo pequeño de informáticos y técnicos que se encargarían de diseñar el proceso de producción y de desarrollar herramientas automáticas para que ésta fuera eficiente, al mismo tiempo que se encargarían de diseñar el sitio Web y los servicios que se presentaría en él. Finalmente, la BVC se abrió a la consulta pública en la Web en julio de 1999 con aproximadamente más de 2.000 obras de más de 400 autores¹¹.

El objetivo principal de la BVC es ya en sí mismo ambicioso: la digitalización de las obras fundamentales del patrimonio cultural español e iberoamericano para poner a disposición de la comunidad científica y de la población de hispanohablantes un total de 30.000 obras. Hoy en día está formada por más de 6000 DLOs¹² que conforman una colección de textos completos de obras clásicas de autores de ambos lados del Atlántico y de todos los tiempos, desde los Cronistas de Indias hasta los contemporáneos. El proyecto incluye obras de renombre y otras menos conocidas o difundidas, con el fin de preservar para el futuro el patrimonio cultural hispano y para proporcionar el acceso a la información y a la cultura a todos los “ciudadanos de la Red”, fundamentalmente a la comunidad de más de cuatrocientos millones de

¹¹ La inauguración de la Biblioteca Virtual Cervantes (27/07/99) fue una noticia reflejada en toda la prensa local y nacional como un hito cultural sin precedentes, surgido a tenor de Internet. Destacamos, algunas de estas noticias: Ana Álvarez. La Universidad de Alicante crea la primera biblioteca virtual en español. *ABC*, 28 de julio de 1999; Laura Cárdenas. www.cervantesvirtual.com. *El Mundo*, 28 de julio de 1999; La primera Biblioteca Virtual de habla hispana. *Cinco días*, 28 de julio de 1999; S. Serrano y J. Esquembre. Una gran biblioteca acoge en Internet 2.000 obras en castellano. *El País*, 28 de julio de 1999; Lucía Enguita. 30.000 clásicos españoles llenarán Internet en los próximos dos años. *El País*, 29 de julio de 1999; etc.

Posteriormente, la BVC ha estado siempre en el punto de mira de los medios, al considerarse un hecho sin parangón para la cultura española en la Sociedad de la Información. *Vid.* p. ej. J.J.M.G. La biblioteca virtual Cervantes registra 92.000 visitas al día. *El País*, 9 de septiembre de 1999; Éxito de la Biblioteca Virtual de Alicante. *El País*, 5 de enero de 2000; La biblioteca virtual Miguel de Cervantes permite la "lectura sonora" de cien obras. *El Mundo*, 16 de febrero de 2000; La Universidad de Alicante inaugura un 'escaparate virtual' para editoriales. *El Mundo*, 12 de abril de 2001; Olga de Nova. La Biblioteca Virtual 'enlaza' con la UNESCO. *El Mundo*, 16 de mayo de 2001; etc.

¹² Concretamente, el 27/07/01 tenía 5.927 obras en formato digital. Nosotros entendemos esas obras como DLOs, es decir, como unidad documental para la asignación de metadatos. Cada DLO puede estar compuesto, como veremos, por múltiples páginas Web, además de por diferentes ficheros de imagen (facsimiles, p. ej.), sonido, etc.

hispanohablantes¹³. José Carlos Rovira, resume el objetivo de la Biblioteca Virtual Miguel de Cervantes, con estas palabras:

*La biblioteca universal era el objetivo último que, centrado en textos en castellano y abierto a las otras lenguas peninsulares, movía nuestra iniciativa originalmente*¹⁴.

Este proyecto, con más de ciento quince millones de peticiones de información en dos años y medio¹⁵, demuestra su utilidad y su universalidad desde el principio. Su carácter universal no se justifica sólo por sus contenidos, sino también por la facilidad y gratuidad del acceso y la condición de servicio público. Este carácter universal de la BVC que advierte el que fuera su director en la época de expansión de la misma, ha sido reconocido también por otros autores. Por ejemplo José Antonio Ontalba¹⁶ en un estudio sobre los proyectos de biblioteca digital que se han gestado en los últimos años en nuestro país, la clasifica como *biblioteca global/universal temática por su intención de dar acceso a toda la literatura española a texto completo*.

¹³ Este objetivo de dar acceso a tod@s los ciudadanos está en consonancia con los principios de las políticas de información internacionales (G7, eEurope, IST del 5PM) y nacionales (InfoXXI) en lo relativo a la creación del acceso universal a la información. Principios que, por otra, hemos defendido y matizado a lo largo de esta investigación en el contexto de las bibliotecas digitales y la normalización de metadatos. *Cfr.* 8.2.1.

¹⁴ José Carlos Rovira Soler, Sobre la experiencia en nuevas tecnologías de la información de la universidad de Alicante y la Biblioteca Virtual Miguel de Cervantes [documento RTF]. En: *Congreso Internacional de la Lengua Española (2. 2001. Valladolid)*. Disponible en: http://cvc.cervantes.es/obref/congresos/valladolid/fronteras/ponencia/rovira_j.doc, p. 3 (consultado el 17 de octubre de 2001). También disponible en: http://cervantesvirtual.com/noticias/congresolengua/ponencia_rovira.htm Esta concepción de la BVC como "biblioteca universal" podría justificar que el nombre formal del proyecto sea "Biblioteca Virtual". *Cfr.* 8.1. *Vid. infr.* Fig. 62.

¹⁵ Las estadísticas concretas de las peticiones, acceso etc. de la BVC, se pueden consultar en: <http://www.cervantesvirtual.com/estadisticas.shtml>

¹⁶ José Antonio Ontalba y Ruipérez. *Op. cit.*, p. 404.

La BVC tiene la vocación y compromiso de continuidad. Intenta promover la suma de esfuerzos y proyectos de otras universidades, instituciones y empresas españolas, latinoamericanas o de cualquier otro lugar, interesadas por el conocimiento, estudio y difusión de la lengua española. Esto ha supuesto desde el principio una convocatoria totalmente abierta para lo que se arbitraron convenios y un ámbito de gestión adecuado, abierto a la colaboración. Así, hoy en día la BVC tiene convenio con más de ochenta instituciones de todo el mundo¹⁷ de las que treinta y nueve son universidades, de entre las que destacan, por ejemplo, las Universidades de Cornell, Harvard y Standford.

El fondo de la BVC es único en el mundo, tanto por la cantidad de obras incluidas en él como por su continuo crecimiento y por la fidelidad de los textos, sometidos a un minucioso proceso de control filológico y de tratamiento automatizado de la información. Actualmente (julio de 2001) ofrece el legado de los más importantes autores de la tradición cultural hispanoamericana, escritos o publicados entre los siglos XII y XIX tanto de prosa, verso, teatro, etc. Así se encuentran digitalizadas obras de Cervantes, Quevedo, Calderón, Lope de Vega, etc., incluso en algunos casos (*v.gr.* Cervantes y Galdós), las obras completas, y en otros, partes de algunos trabajos contemporáneos a través de Primeravista¹⁸.

¹⁷ El listado completo de las entidades que colaboran con el proyecto se puede consultar en: <http://www.cervantesvirtual.com/asociados/asociados.shtml> La Universidad Carlos III de Madrid es una de las universidades que firmó un convenio de colaboración con la Biblioteca Virtual Cervantes. *Acuerdo de colaboración entre la Universidad de Alicante y la Universidad Carlos III en el proyecto Biblioteca Virtual Miguel de Cervantes*. Firmado por D. Andrés Pedreño y D. Gregorio Peces-Barba, en Alicante el 12 de julio de 1999, [5 p.]. Entre otras instituciones que mantienen convenios con la BVC, merece destacar que recientemente (22/11/01), Luis Racionero, director de la Biblioteca Nacional, acordó firmar un convenio con la BVC para que se digitalicen algunos fondos de la BNE, que hasta ahora no había cedido ninguno de sus valiosos fondos al proyecto.

¹⁸ Primera Vista <<http://www.primeravistalibros.com>> es un proyecto creado y mantenido por la BVC e inaugurado el 4 de junio de 2001 en el círculo de Bellas Artes de Madrid. En él se plantea un proyecto de cooperación con el gremio de editores (más de 30 editoriales) para la digitalización y acceso a fragmentos de obras de autores contemporáneos. Además este proyecto permite dotar de contenidos actuales a la BVC sin contravenir los intereses de la industria editorial.

Además de un catálogo completo del fondo, del que tendremos oportunidad de hablar desde el punto de vista técnico y bibliotecario¹⁹, la BVC presenta una serie de secciones que se denominan, dentro de ella, "*portales*", dedicando un espacio específico y monográfico, bien a distintas instituciones²⁰ que han contribuido con sus valiosos fondos para que se digitalizaran y formasen parte de la colección de textos digitales de la BVC, bien a temas específicos que resaltan el carácter multitemático y multicultural de la biblioteca Virtual Cervantes²¹. Otra forma de seccionar la vasta cobertura de los fondos de la BVC es a través de lo que se denominan *bibliotecas de autor* que son "pequeñas bibliotecas" dentro de ella, dedicadas a autores clásicos o contemporáneos donde se describe la biografía de esos autores con enlaces a sus obras digitalizadas en la BVC²² y otros recursos sobre tales personajes de la cultura hispánica. Todas estos "portales temáticos" y bibliotecas de autor, etc. suponen la versión digital de la biblioteca entendida como colección de obras sobre un mismo tema o de un mismo autor, muy próxima del concepto de serie bibliográfica en el

¹⁹ Vid. *infr.* 9.2.2. y 9.3.2.

²⁰ Algunos de estos portales institucionales son, entre otros:

- La Biblioteca Nacional de Argentina: <http://www.cervantesvirtual.com/portal/BNA>
- El Colegio de México: <http://www.cervantesvirtual.com/portal/ECM/index.html>
- La Caja de Ahorros del Mediterráneo: <http://www.cervantesvirtual.com/portal/CAM>

²¹ Como ejemplos de los portales temáticos podemos mencionar:

- Portal DOXA de Filosofía del Derecho: <http://www.cervantesvirtual.com/portal/DOXA>
- Portal de Literatura paraguaya: <http://www.cervantesvirtual.com/portal/paraguay>
- La Biblioteca infantil y juvenil: <http://www.cervantesvirtual.com/portal/platero>

Dentro de las divisiones temáticas que contempla esta biblioteca digital es preciso destacar la Biblioteca de Historia <<http://www.cervantesvirtual.com/historia>> que, aunque no se incluye dentro de los portales temáticos es uno de ellos con gran riqueza de contenidos que secciona a su vez, importantes portales como el de las constituciones hispanoamericanas, la monarquía Hispánica, los jesuitas, el tiempo y la humanidad, o el portal de Carlos V.

²² En noviembre de 2001, la BVC contaba con: 16 bibliotecas dedicadas a autores clásicos, como por ejemplo, la de Calderón de la Barca <http://www.cervantesvirtual.com/bib_autor/Calderon> que contiene sus obras completas; 15 bibliotecas de autores contemporáneos, como por ejemplo la de Mario Benedetti <http://www.cervantesvirtual.com/bib_autor/mbenedetti>; 3 bibliotecas de autores y grupos teatrales (Dagoll Dagom, Pepe Rubianes y Tricicle), a las que se une además una nueva sección dedicada a cineastas inaugurada con la biblioteca de Rafael Azcona <http://www.cervantesvirtual.com/bib_autor/Azcona/montaje.html>.

mundo tradicional impreso (v. gr. *colección de autores castellanos, biblioteca básica de economía o biblioteca de Rafael Alberti*).

También tiene otras secciones con catálogos individualizados del catálogo general de la BVC, donde podemos subrayar, por ejemplo, la sección hemeroteca²³, que pretenden digitalizar revistas científicas de diferentes áreas (por ejemplo *Doxa* (filosofía del derecho) o *América sin nombre*), o el catálogo de tesis doctorales²⁴. Además de estas secciones de componente eminentemente textual, destacan otras, formadas por objetos de información digital y multimedia que también forman parte de la colección de la BVC y que aparecen agrupados en las siguientes secciones: *la biblioteca de imágenes, la biblioteca de voces y la biblioteca de signos*.

Además de la riqueza de su colección digitalizada, esta biblioteca tiene la vocación de ser un foro de comunicación, e incluso un portal de servicios. La denominación de "portal" está muy presente en la BVC ya que incluye otros servicios de valor añadido además de lo estrictamente bibliográfico. No obstante, a pesar de que ya hemos comentado la popularidad de este término, desde nuestro punto de vista el concepto de portal tiene una connotación empresarial y económica (asociada sobre todo a los grandes portales horizontales), además de que, en el ámbito de las bibliotecas digitales, el término portal se asocia o se vincula más a las bibliotecas híbridas (y ésta es enteramente digital). A tenor de estas reflexiones, sería más conveniente pues, decir que la BVC es además una *comunidad virtual* de usuarios con intereses culturales, según el concepto de comunidad virtual que defienden

²³ Para ampliar esta sección de *Hemeroteca* dentro de su fondo digital, la BVC ha suscrito un convenio con el Proyecto Highwire Press <<http://intl.highwire.org>> de la Universidad de Stanford. Recordemos que la Universidad de Stanford es una de las universidades americanas protagonistas de la DLI (*Digital Library Initiative*). Cfr. 8.2.1.

²⁴ En este catálogo de tesis <http://www.cervantesvirtual.com/tesis/tesis_catalogo.shtml> se encuentran digitalizadas más de 500 tesis de Universidades de todo el mundo. Además de los accesos por índice alfabético de autor y CDU, común a todos los catálogos, presentan también índices con hipervínculos que facilitan el acceso por Universidades y por la clasificación de códigos UNESCO. Destaca sin embargo, que en esta sección no se incluye la búsqueda por palabras clave. *Vid. infr.*

Sánchez y Saorín²⁵, donde no sólo encuentran servicios de valor añadido sino también un lugar donde interactuar entre los miembros que la componen. Dentro de esos servicios, que le dan el carácter de portal destacan verbigracia, los foros de debate, la comunicación con el bibliotecario²⁶, el envío de tarjetas, el boletín de noticias Dulcinea o el curioso e interesante servicio de "Trueque" donde se interconectan a los usuarios para encontrar obras o ejemplares difíciles de conseguir. Incluso tiene también la opción de personalización del acceso (característica habitual en los portales convencionales) a la BVC, gracias al servicio "mi biblioteca" donde el usuario, una vez suscrito puede, entre otras cosas, "señalar" digitalmente su lectura en los libros digitales que elija.

El proyecto Biblioteca Virtual *Miguel de Cervantes* conlleva otras implicaciones que maximizan su importancia y que nos han llevado a elegirla como foco de nuestra planificación de la aplicación de metadatos. Siguiendo la información suministrada por la propia BVC²⁷, sintetizamos a continuación algunas de ellas:

- Constituye una línea activa de investigación a través de proyectos (actualmente en tramitación) que se encuadran en la filosofía del V Programa Marco de la UE.
- Representa una invitación para todos los portales culturales y bibliográficos, como punto de acceso continuamente actualizado y en permanente innovación en materia de edición digital, bases de datos, software de edición y tratamiento de texto, imagen y sonido en general.

²⁵ M^a Vanessa Sánchez Arce, Tomás Saorín Pérez. Las comunidades virtuales y los portales como escenarios de gestión documental y difusión de información. *Anales de Documentación*, 2001, n° 4, p. 215-228.

²⁶ Cfr. Capítulo 8, nota 20.

²⁷ *La Biblioteca Virtual Miguel de Cervantes* [documento HTML]. Alicante: Biblioteca Virtual Miguel de Cervantes, Disponible en: <http://www.cervantesvirtual.com/proyectoES/BIMICESA03.shtml> (consultado el 17 de julio de 2001).

- Se realiza la edición de textos actuales, educativos o científicos, cuyos autores aspiren a alcanzar la máxima difusión.
- Existe una política de desarrollo progresivo de los recursos potenciales existentes dentro de ella.

La gran calidad del resultado final se ha conseguido con la aplicación de nuevos métodos de edición electrónica para grandes cantidades de información documental y con el máximo rigor en el tratamiento de los textos. Se trata de un proyecto abierto, que aspira a convertirse en uno de las más completas bibliotecas digitales en español, en el principal referente internacional para el estudio de textos científicos y de la cultura hispana e incluso, en un "portal" de índole cultural para el ámbito de la lengua castellana. Lejos de ser una colección estática de libros digitalizados, la BVC tiende a ser un vehículo para la comunidad hispánica y una "ventana" a la literatura y la cultura para todos aquellos investigadores en las lenguas hispánicas, o de forma categórica, parafraseando a Alejandro Bía y Andrés Pedreño²⁸, la Biblioteca Virtual *Miguel de Cervantes*, pretende ser *la voz hispánica en la Web*.

Finalmente debemos hacer una matización en consonancia con la determinación del concepto de "biblioteca virtual" que hacíamos en el capítulo anterior. La BVC no es una biblioteca virtual en el sentido técnico de la palabra, es decir, no trata de simular los espacios de la realidad física de una biblioteca a través de mundos virtuales en VRML o en X3D. Tampoco podemos decir que se adapte a la tendencia más consensuada de denominar "biblioteca virtual" a aquella que no alberga físicamente el fondo, sino que se basa en la descripción y organización de una colección virtual. Sin embargo, el nombre formal de "biblioteca virtual" que se le ha dado a esta proyecto, tremendamente acertado, si partimos de la doble visión de

²⁸ Alejandro Bía, Andrés Pedreño. *Op. cit.*, p. 149.

Saorín que explicábamos en el capítulo 8 y que recogemos en la Fig. 62²⁹, a continuación.

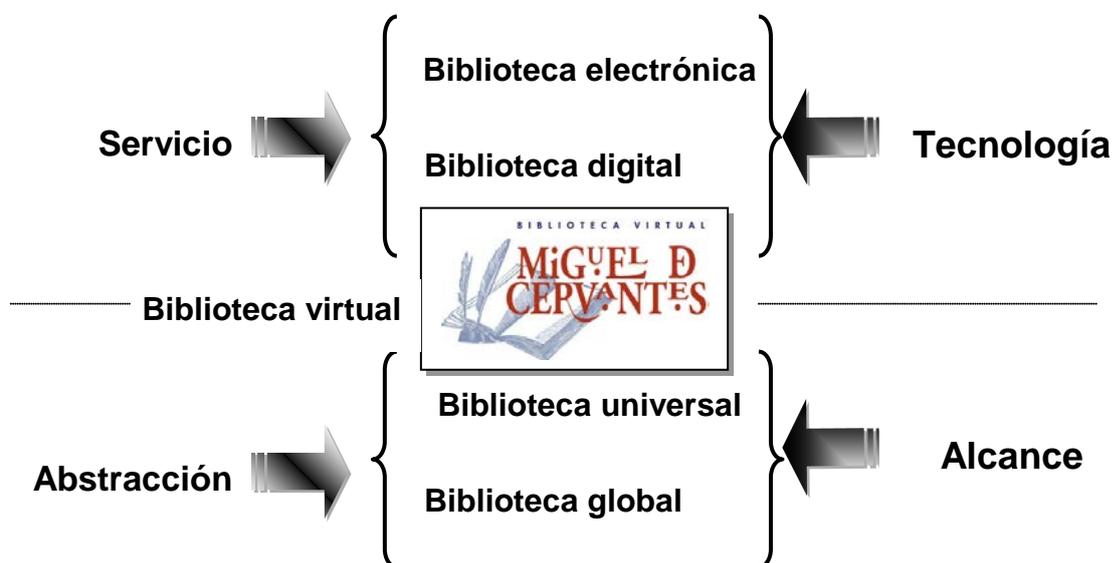


Fig. 62. Expresión gráfica del concepto de *biblioteca virtual* aplicado a la BVC

En la Biblioteca Virtual *Miguel de Cervantes* constituye, por un lado, la materialización del concepto de biblioteca asociado a la organización de un servicio, donde las colecciones se encuentran accesibles de modo unificado a través de la tecnología; por otra parte, su alcance y la definición abstracta de su filosofía hacen de ella una biblioteca de carácter global con una proyección mundial de sus servicios, que hacen pensar en la biblioteca universal borgiana o en la biblioteca de Alejandría.

9.2.2. La BVC como creadora y gestora de información digital

La Biblioteca Virtual *Miguel de Cervantes* se ha convertido en poco tiempo en el referente básico para el acceso digital a la cultura española. La importancia, envergadura, rigor y la calidad de este proyecto nos lleva a ser partícipes en el

²⁹ Imagen basada en: Tomás Saorín. *Los Portales Bibliotecarios...* Op. cit., <http://gti1.edu.um.es:8080/portales/FGSR-portales-Bibliotecas.PDF> Cfr. 8.1.

dinamismo tecnológico que tienen que enfrentar este tipo de servicios de información Web. Además de estas características, y de otras que se puede colegir de la importancia del proyecto expresada en el apartado anterior, existen otros aspectos de índole técnico y bibliotecario que debemos considerar antes de plantear la aplicación de metadatos en bibliotecas digitales, y que recopilamos a continuación.

La BVC es una biblioteca productora y poseedora de su colección digital, ya que se encarga tanto de la digitalización de los textos en formato convencional y de su almacenamiento, como de su adaptación y publicación en la Web, y al mismo tiempo también organiza, gestiona y da acceso a la colección que crea o edita digitalmente. Esto no hace que cambien en ella los objetivos y funciones tradicionales de la biblioteca, conservar y difundir, ni las tareas de formación, (selección, adquisición), organización y difusión de la colección. Sin embargo, el hecho de que la biblioteca sea editora del fondo hace que las tareas básicas de la biblioteca cobren otra dimensión y se añadan otras nuevas en un complejo sistema informativo, donde se altera la cadena documental, en los aspectos que reflejamos a continuación³⁰:

- Formación de la colección. La selección se realiza sobre las obras impresas que se digitalizan. En el caso de la BVC se seleccionan obras de la literatura hispánica cuyo texto original haya sido cedido por la institución que las custodia, por el propio autor o por acuerdo con la editorial que las edita, en el caso de autores contemporáneos, o bien que hayan pasado los ochenta años que la Ley marca en España desde el fallecimiento de un autor para que sus textos pasen a ser de dominio público.

De esta forma, la adquisición, entendida en el sentido tradicional, queda reducida a la cesión, normalmente temporal, de la obra original para la digitalización. No

³⁰ *Cfr.* 8.2.1.

obstante, para nosotros, la obra se puede considerar "adquirida" desde el momento en que se almacena en un servidor de la BVC en formato digital.

- La creación del DLO implica convertir un documento tradicional (texto, vídeo y audio analógico, etc.) en un objeto de información electrónica. Es esta nueva dimensión del trabajo en la BVC, la que la convierte en editora del documento digital.
- La organización de los DLOs, que en el mundo bibliotecario tradicional implica el análisis documental y la ordenación de los fondos, en el contexto bibliotecario digital debe conllevar la creación gestión de metadatos.

Es este el aspecto que más nos interesa destacar, pues la BVC no utiliza un sistema de organización y recuperación de la información basado en metadatos en sentido estricto, tal y como defendemos en esta tesis doctoral, donde el fundamento y objeto de la metainformación es el documento entendido como objeto (DLO) que es, por otra parte el que verdaderamente forma parte de la colección digital. Basa la organización posterior recuperación de su colección en una breve catalogación del documento impreso.

- Por último, la difusión que en la BVC, comprende: la propia edición (ya que en una biblioteca digital la difusión de la colección está mediatizada por la edición electrónica de los documentos seleccionados), como la recuperación de información y el resto de servicios que se incluyen para difundir las obras digitalizadas.

Con todas estas consideraciones generales, en la BVC, como en el resto de bibliotecas digitales, no podemos hablar de "cadena documental" sino de una generación de flujos de trabajo para la producción, organización y recuperación de DLOs. Teniendo en cuenta esto, en la BVC se distinguen dos tipos fundamentales de producción documental:

- 1) Portales especializados (bibliotecas de autor y portales temáticos o especializados, como la biblioteca de signos o la biblioteca de voces) que contienen información multimedia (grabaciones sonoras, vídeos digitales de los autores, etc. o por ejemplo, en la biblioteca de signos, vídeos que recogen las imágenes de la representación de un libro en el lenguaje de signos para sordomudos)
- 2) La producción masiva de libros³¹, DLOs textuales, sin elementos multimedia, editados de una forma homogénea y uniforme.

Es importan resaltar que en la BVC, el objetivo de *producción masiva* de documentos digitales (recordemos que el proyecto espera digitalizar y poner al acceso público en la Web 30.000 documentos), debe tenerse en cuenta a la hora de tomar cualquier tipo de decisión relacionada con los metadatos, al mismo tiempo que sirve para justificar la necesidad de automatización de procesos (el marcado asistido por ordenador o la transformación automática de XML a HTML y, en nuestro caso, la previsión de extracción-asignación automática de metainformación, son absolutamente necesarios).

El proceso de producción del libro digital es el aspecto fundamental de esta biblioteca. En el proceso de producción marca una visión especial de la *cadena documental* aplicada a bibliotecas digitales productoras y poseedoras de su colección,

³¹ Los DLOs que surgen de este tipo de producción digital se denominan, "libros digitales" (para la BVC, para nosotros, DLOs o unidades documentales electrónicas en la biblioteca digital). Según Bia y Pedreño en el primer año de la biblioteca, la creación de estos "libros digitales" supuso 54.936 ficheros HTML, de los cuales 38.354 eran capítulos o divisiones estructurales de los libros, 11.835 imágenes, y el resto eran páginas de cubierta y ficheros de notas al pie. Alejandro Bia, Andrés Pedreño. *Op. cit.*, p. 156. Esta distinción de los ficheros nos sirve para reflexionar sobre el nivel de detalle (*granularity*) que deberán tener las descripciones de metadatos, es decir si se asignan metadatos para cada uno de los objetos de información que conforman un "libro digital" o si por el contrario se establece la unidad de DLO a nivel "libro" o conjunto de ficheros que conforman la versión digital, publicada en la Web, del correspondiente libro impreso, lo que correspondería al *nivel de objeto primario* en la diferenciación que establece la LC. *Cfr.* 3.3., Capítulo 3, nota 70.

donde, como vimos³², se establece un subsistema editorial dentro de la visión sistémica de la biblioteca. En la Biblioteca Virtual *Miguel de Cervantes*, tal y como se concibe en la actualidad, podemos distinguir, al menos, seis fases que se representan en la Fig. 63³³ y que explicaremos a continuación.

³² Vid. Capítulo 8, Fig. 60 y 8.2.1.

³³ Basado en la presentación un póster en la Conferencia Internacional de la ACH/ALL: Alejandro Bia. Technical Aspects of the Production Process of Digital Books Using XML-TEI at the Miguel de Cervantes Digital Library [documento HTML]. En: *International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing (2001. New York)*. New York: University, Humanities Computing Group, rev. 15 de mayo de 2001. Disponible en: http://www.nyu.edu/its/humanities/ach_allc2001/posters/bia/index.html (consultado el 30 de mayo de 2001) y en: Alejandro Bia, Andrés Pedreño. *Op. cit.*, p. 157

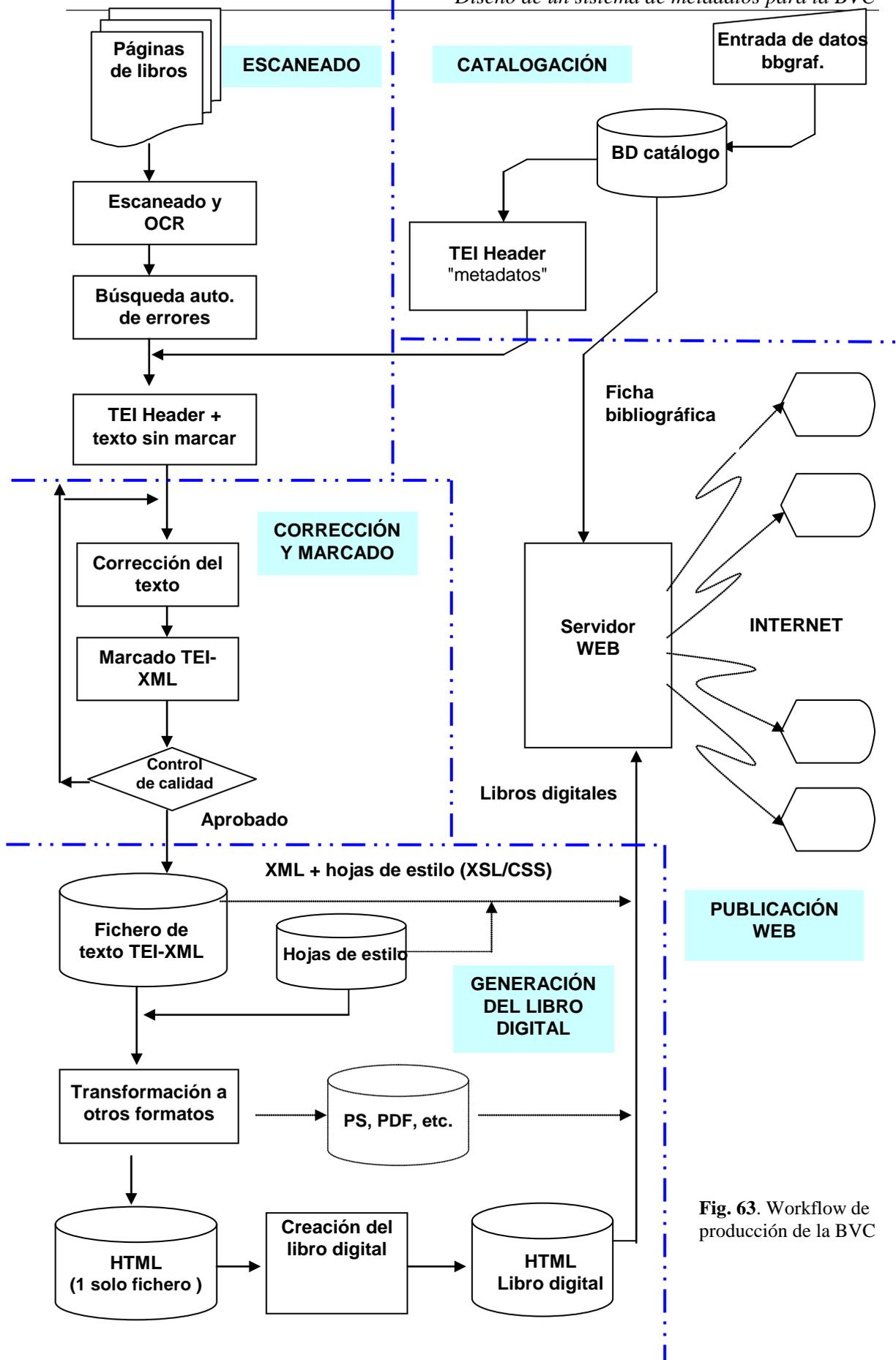


Fig. 63. Workflow de producción de la BVC

Para replantear el modelo de biblioteca digital basado en metadatos, es necesario analizar el *workflow* de la creación de los documentos digitales en la BVC, cuya representación acabamos de ver, distinguiendo las siguientes fases en los flujos de trabajo de producción:

1. Búsqueda bibliográfica y selección. Esta fase implica la localización de libros interesantes desde el punto de vista de los objetivos de la BVC, que posteriormente se digitalizarán y formarán parte de su fondo. Una vez que se ubican dichos materiales bibliográficos en soporte tradicional, los bibliotecarios, de forma semejante a como se realiza la adquisición en una biblioteca convencional realizan las tramitaciones oportunas para conseguir el libro impreso. La "adquisición", como anticipamos, se traduce normalmente, en el contexto de la BVC, en conseguir el préstamo o cesión de una obra determinada. Desde ese momento, se introducen los datos bibliográficos elementales (basados en campos MARC, aunque no en formato MARC) de esa publicación y se almacenan en una base de datos en Access.

Esta información bibliográfica que almacenan en Access tiene diversas utilidades en la BVC: ayudar a controlar el proceso de producción y el de publicación, es la base de datos sobre la que se basan las búsquedas en la BVC y es la fuente tanto para la creación de la cabecera TEI mínima que se asociará al texto sin marcar, como para la producción automática de "fichas de catálogo" accesibles en Internet como la que muestra el ejemplo de la Fig. 64³⁴.

2. "Catalogación". Cuando se recibe el libro, de igual forma que en las bibliotecas tradicionales, se realiza una catalogación completa pero no exhaustiva (siguiendo

³⁴ Nótese la excesiva vinculación a la biblioteca tradicional. En la propuesta que realizamos aquí pretendemos dejar patente la correlación que tácitamente defendemos en esta tesis: nuevos sistemas de información-nuevas formas de representación y recuperación de la información. Desde nuestro punto de vista, en vez de una ficha de catálogo convencional deberían visualizarse los metadatos asociados al DLO (*Cfr. infr.* Fig. 67). Ya que la información que reportan estas fichas es sobre el libro tradicional impreso que se cataloga en la fase que describimos a continuación y no sobre el DLO.

campos MARC³⁵, en la base de datos en Access). Se dan valores a los campos de título autor, etc. así como materias a partir de un Tesauro de Literatura y la Lista de encabezamientos para Bibliotecas Universitarias de la Universidad de Sevilla; en el caso de la biblioteca Joan Lluís Vives, o de los fondos de la Biblioteca de Cataluña, se asignan encabezamientos en catalán de la Universidad de Valencia. Además se establece una nota (campo 500 en MARC) donde se pone la referencia bibliográfica completa de la obra digitalizada³⁶. El resultado de la catalogación será lo que visualicen los usuarios en Internet como fichas asociadas a una obra (Fig. 64) y generadas automáticamente a partir de la base de datos en Access. Además, en el momento de catalogación se le asigna un número (como si se tratase de un número de registro convencional) que es el que identifica permanentemente a la obra durante todo el proceso de producción y edición y que será el código utilizado en todos los ficheros relacionados con el libro. En la obra que estamos tomando como ejemplo, *Las esmeraldas* de Joaquín Dicenta, tiene el número 1179 en el catálogo, por ello, la ficha en el "catálogo de obras" (Fig. 64) tiene esta ubicación: <http://www.cervantesvirtual.com/FichaObra.html?Ref=1179>, de igual forma que el fichero marcado en TEI-XML tiene la denominación: 001179.XML.

³⁵ Es importante destacar que se introduce la información basándose en campos MARC (fuente de acceso principal, autor, título, edición, publicación), pero NO en formato MARC, es decir, no en formato de codificación compatible ISO 2709. De esta forma, en el planteamiento de asignación de metadatos que proponemos en este capítulo, no podemos plantear el establecimiento de *crosswalk*, ni la utilización de Z39.50 para la recuperación que evitarían una "conversión retrospectiva". Cfr. 8.3.2., Vid. Glosario, *crosswalks/crossroads*.

³⁶ La relación entre el libro original y el libro electrónico no tiene por qué ser uno a uno, es decir, un libro impreso puede dar lugar a varios "libros electrónicos" (a varios DLOs), verbigracia en el caso de que aquel sea una colección de obras completas o similar. La ficha que se muestra en la Fig. 64 es un ejemplo de esto pues, como consigna la nota, la obra *Las esmeraldas* de Joaquín Dicenta, está comprendida en las páginas 9-83, del libro titulado *Novelas*.



Fig. 64.

Ficha de catálogo de obras de la BVC:
Las Esmeraldas de J. Dicenta (Web)

3. *Escaneado* de documentos. Desde el momento de la catalogación del documento comienza el proceso de digitalización del mismo produciéndose dos tipos de ficheros básicos: imágenes escaneadas y documentos textuales a los que se les pasa un OCR. Se realiza una corrección automática del texto de tal forma que si se detectan muchos errores se ajusta el escáner y se vuelve a digitalizar y sino, pasa a la siguiente fase de corrección y marcado de documentos. En este momento se incorpora la cabecera TEI extraída de la base de datos catalográfica en Access, al documento digitalizado³⁷.
4. Corrección y Marcado. Esta fase implica dos procesos muy importantes en el resultado final de la BVC:
 - a) Por un lado, la corrección cualificada por filólogos y especialistas en literatura, historia, lingüística, etc. en un editor de textos (no sólo errores del OCR, sino también de la edición base, contrastando el texto con otras ediciones).
 - b) Por otro, el marcado del documento, esto es, aplicar las marcas que describen la estructura del documento y que indican cómo se va a construir el libro

³⁷ Vid. 5.2.2.1. La TEIH, del documento que estamos utilizando como referencia y ejemplo (001179: *Las esmeraldas* de Joaquín Dicenta) es la que utilizamos en Ejemplo TEI Header (2) del apartado 5.2.2.1. al hablar de las cabeceras TEI.

digital posteriormente. El marcado se realizará a partir de una DTD adaptada de TEI-XML (P4) (CERVANTE.DTD³⁸) donde se establecen elementos y atributos relativos a la fragmentación de capítulos, índice del libro, etc. con la ayuda del software XMetal³⁹.

Antes de pasar a la siguiente fase se valida el documento y se hacen las últimas correcciones. A partir de este momento, todo el proceso de generación del "libro digital" y de su publicación en la WWW es automático.

5. Generación del "libro digital". Antes de la publicación definitiva en la Web se transforman los documentos en TEI-XML a los formatos habituales de visualización en la WWW, esto es, HTML y PDF fundamentalmente. El

³⁸ Algunos aspectos importantes de la DTD de la BVC son por ejemplo: la normalización de los valores de los atributos cambiando las declaraciones CDATA por una lista fija de valores, en aras a evitar posibles errores de marcado y para forzar el uso de valores que serán fundamentales en el proceso siguiente. Así por ejemplo el elemento `cit` se muestra así en `cervante.dtd`:

```
<!ELEMENT cit
  (q | bibl | biblFull | ptr | ref | xptr | xref | index |
  interp | interpGrp | lb | milestone | pb | gap | anchor)+ >
```

Otro aspecto interesante es que la DTD fuerza algunas decisiones sobre el marcado por ejemplo las divisiones numeradas, y también la identificación de objetos textuales (poemas, etc.) utilizando valores específicos para elegir. P. ej.

```
<!ATTLIST div0
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  decls IDREFS #IMPLIED
  type (work | book | part | chapter | act | scene | tale |
letter | introduction | preliminaries | dedication | prologue |
personae | coverpage | toc | epilogue | colophon) #IMPLIED
  org (composite | uniform) "uniform"
  sample (initial | medial | final | unknown | complete)
"complete"
  part (Y | N | I | M | F) "N"
  TEIform CDATA "div0" >
```

³⁹ Uno de los objetivos es automatizar al máximo el marcado ya que los OCRs actuales sólo pueden reproducir el marcado de procedimiento o formato de un documento, pero se necesita añadir además a los textos un marcado descriptivo y/o estructural. Por ello, en la BVC se han construido *parsers* para convertir documentos, a partir de un editor de textos, a TEI-XML. Estos *parsers* se basan en el uso del estilo para localizar encabezamientos, deducir bordes de división, añadir marcas de párrafo `<p>`, texto `<text>`, fuente `` y cuerpo `<body>`. Esto genera un marcado básico sobre el que se realizarán correcciones.

documento marcado pasa por diferentes *parsers* de transformación basados en XSLT (*eXtensible Style Sheet Language Transformation*): el primero de ellos, desarrollado por el equipo de la BVC y denominado MakeBook transforma el texto marcado TEI en un fichero HTML, que luego se dividirá en partes estructurales (capítulos, etc.), luego se le añadirá la tabla de contenido hipertextual y los botones de navegación y demás elementos contenidos en una plantilla aplicada para que todos los libros tengan el mismo aspecto. De igual forma, para la creación de libros en PDF se utiliza el *parser* de transformación XT⁴⁰ basado también en XSLT. Existen además producciones digitales especiales como la publicación de manuscritos donde también se aplica XSLT para construir libros digitales que alberguen enlaces e imágenes y que se puedan visualizar en la Web⁴¹.

6. Publicación Web. Finalmente, fruto de la transformación a HTML a través de XSLT, los libros digitales (los documentos entendidos como objetos de información, en el sentido que defendemos en esta tesis) se pueden publicar y son accesibles en el sitio Web de la BVC.

El sistema de recuperación aplicado a la colección de documentos digitalizados y publicados en el servidor de la BVC se basa fundamentalmente en un catálogo⁴² (Figs.65 y 66) o en la búsqueda rápida que aparece en la página principal:

- La búsqueda en el catálogo, como muestra la Fig. 65, se basa en un índice alfabético, al estilo de los viejos catálogos manuales, de autores y títulos (o

⁴⁰ XT es una implementación en Java XSLT, de James Clark's que también permite trabajar con plantillas. Más información sobre este software: <http://www.jclark.com/xml/xt.html> (última versión 05/11/1999).

⁴¹ La producción de manuscritos y ediciones facsimilares en formato digital se detalla en: Alejandro Bia, Manuel Sánchez. A reusable hypermedia design for DL manuscripts, based on XML, XSLT and Java. (ejemplar original de autor).

⁴² Catálogo de la BVC: <http://cervantesvirtual.com/catalogo.shtml>

primeros versos en el caso de la poesía) y CDU. El acceso es simplemente a través de la navegación dirigida alfabéticamente que conduce a un listado hipertextual de autores o de títulos a los que aparece asociada un icono de acceso a la ficha (ej. Fig. 64), así como otras características de la obra (por ejemplo, un icono que evoca que pertenece a la biblioteca de imágenes, o el icono que muestra que dicha obra tiene asociado un índice de concordancias⁴³). En estos listados aparecen en diferente color los enlaces a tesis, de los enlaces a obras propiamente dichas y de los *links* a estudios críticos sobre una obra, asimismo se destacan, en color gris, las obras que aún están en proceso de edición o corrección. Además de los índices, presenta una interfaz de búsqueda por palabra clave en los campos de autor, título, materia o período temporal de la obra, las búsquedas se realizarán, como hemos dicho, en la base de datos bibliográfica en Access.



Fig. 65. Recuperación de información y catálogo en la BVC



Fig. 66. Catálogo de títulos en la BVC

- La búsqueda rápida permite la recuperación de información por palabra clave restringiendo la búsqueda a todo el catálogo, a la información de los portales, al contenido de los libros o a las dos últimas. La búsqueda en el catálogo se realiza igualmente sobre los campos buscables (autor, título y materia) de la base de

⁴³ Los índices de concordancias asociados a determinadas obras son una de las opciones de búsqueda más características y particulares que tiene algunas de las obras de la BVC. Estos índices se basan en el sistema TACT desarrollado por Bradley. Vid. Text Analysis Computing Tools (TACT): <http://www.chass.utoronto.ca/cch/tact.html>

datos en Access. La recuperación de información en los portales o en el contenido de los libros o en ambos, se basa en la indización automática del texto completo sobre los contenidos de todo el sitio Web realizada a través del motor de búsqueda HtDig⁴⁴. Por último, la opción de "búsqueda avanzada" remite a la interfaz de búsqueda del catálogo, señalado con una elipse en la Fig.65.

A raíz de las pruebas que hemos hecho podemos decir que todas estas búsquedas son eficaces pero insuficientes. Insuficientes en el sentido permiten explotar poco las potencialidades de la rica (desde el punto de vista del contenido) y bien estructurada (desde el punto de vista de la concepción del marcado XML) información que alberga esta biblioteca digital.

9.3. Propuesta de aplicación de metadatos en la BVC

En este apartado haremos constar un diseño de evaluación, planificación e implementación de metadatos, tratando de adaptarlo a las dimensiones del proyecto. Para ello partimos de que la aplicación de metadatos es un servicio y una tarea emergente en las bibliotecas digitales. Proporcionar un acceso intelectual a cualquier tipo de recurso de información, tanto en sentido tradicional, como en las nuevas bibliotecas digitales, implica que "catalogadores", indizadores o creadores de metadatos entiendan las necesidades de los usuarios, relativa a los datos que se crean. Esto es especialmente difícil en las bibliotecas digitales teniendo en cuenta que, aunque el desarrollo del servicio tenga una planificación local, su proyección a través de la red es global, y en esa globalidad se incluye cualquier tipo de usuario, de cualquier condición física⁴⁵, idioma, etc.

⁴⁴ Cfr. 7.2. Más información sobre este software: <http://htdig.sourceforge.net>

⁴⁵ El hecho de mencionar aquí la condición física de los usuarios potenciales tiene que ver con la realidad de la BVC de no excluir a personas con deficiencias visuales y/o audibles ya que desde el principio incluye una "biblioteca de voces" para los usuarios invidentes y una "biblioteca de signos" multimedia para los usuarios sordos. Esta circunstancia es destacable también en lo relativo a la [cont.]

Por otra parte creemos que la planificación de ese acceso intelectual al que hacemos referencia, a través del cual los usuarios podrán recuperar y utilizar los recursos digitales, debe realizarse al principio del proyecto, antes de la puesta en marcha del servicio de información digital. Sin embargo, en nuestro caso, proponemos la aplicación de metadatos en un momento en que la biblioteca ya está funcionando y es una realidad, con lo cual no son posibles planteamientos ideales o teóricos y, en cierto sentido tenemos que hablar de reconversión, o al menos de replanteamiento y reformulación del sistema de RI.

9.3.1. Justificación y objetivos

A lo largo del punto 9.2. hemos reflejado la magnitud y excelencia de la BVC desde el punto de vista de la calidad de los contenidos literarios, de la concepción de biblioteca como comunidad virtual de usuarios (o incluso, como portal cultural) y del desarrollo informático para la producción masiva de documentos electrónicos, libros digitales, o en nuestra concepción, DLOs. Sin embargo creemos que desde el punto de vista bibliotecario el sistema es débil y demasiado vinculado a la idea de biblioteca tradicional en cuanto a los procesos. Por ello, consideramos que es necesario adaptar la organización y la recuperación de la información en la BVC a las últimas tendencias de metadatos, de tal forma que sea —como lo es a nivel filológico-literario e informático— ejemplar desde el punto de vista bibliotecario y de la gestión de la información, ante el nuevo paradigma en el tratamiento de la información que exigen las bibliotecas digitales.

Como se reconoce en todos los estudios profundos de creación y gestión de nuevos sistemas de información electrónica, y como hemos dejado patente en los capítulos precedentes de este trabajo, la búsqueda y recuperación de información es

aplicación de metadatos, ya que por ejemplo, en las pautas para la accesibilidad del contenido Web [WCAG] del W3C reconocen, con un nivel de prioridad 2 la inclusión de metadatos para incrementar la accesibilidad. *Vid.* WCAG: <http://www.w3.org/TR/WCAG10> *Cfr. supr.* Nota 5.

uno de los problemas técnicos más difíciles a la hora de potenciar el rendimiento de una biblioteca digital basada en estructuras descriptivas SGML o XML⁴⁶.

El objetivo es pues, mejorar las posibilidades de búsqueda basándola en la creación de metadatos *a priori*, en el momento de producción del DLO. Consideramos que la BVC debe asumir los estándares de metadatos existentes en la comunidad internacional para mejorar el acceso global futuro y la interoperabilidad en el contexto de la Sociedad de la Información. Para ello sistematizamos a continuación, algunas de las reflexiones que hemos ido soslayando en torno a la realidad actual de la Biblioteca Virtual *Miguel de Cervantes*, haciendo hincapié en la organización y gestión de metainformación y en la recuperación:

- Un proyecto de aplicación de metadatos en bibliotecas digitales, no es un proceso que se pueda concebir aisladamente, sino que requiere una serie de decisiones o fases previas. Cuestiones del tipo: qué documentos conformarán la biblioteca digital, qué calidad que deben tener los materiales seleccionados⁴⁷, a quién va dirigida la biblioteca digital, etc. todas ellas resueltas ya en el planteamiento actual de la BVC.

⁴⁶ Por ejemplo, el archivo Rosetti <<http://www.iath.virginia.edu/rossetti>> menciona la resolución de este problema a través de la base de datos comercial Tamino que ha implementado la norma del W3C XPath y que implementará también XQuery tan pronto como se apruebe el estándar. Stephen G. Nichols, Abby Smith. *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections*. Washington: CLIR, November 2001, p. 69. También disponible en Internet en: <http://www.clir.org/pubs/reports/pub103/pub103.pdf> Afirmaciones como las que se describen en este informe demuestran varias de las reflexiones que apuntamos en esta tesis doctoral: por un lado, que estamos en un proceso de transición en la gestión de la información Web y por otro, que las recomendaciones del consorcio Web se convierten, a pesar de su condición de PAS, en normas *de facto* formales como las definíamos en el capítulo 6. RDF al ser ya recomendación desde el año 1999, se va a implementar de forma creciente por todas las aplicaciones de gestión de información XML.

⁴⁷ En este caso los criterios de selección se aplican sobre el contenido o la calidad literaria y/o científica de los documentos convencionales que se decide digitalizar, no son aplicables por tanto criterios de selección de documentos Web como en el caso de las *subject gateways*. *Vid. supr.* 9.2.2. *Cfr.* 8.2.2.

- Lo que en la BVC es un "libro digital" para nosotros es un objeto (un documento digital entendido como objeto, DLO) y la información sobre él (los metadatos) es también un objeto de información.
- Lo que realmente constituye la colección de una biblioteca digital que es productora y gestora de la misma son los documentos digitalizados, marcados y almacenados en su servidor/es locales esto es los DLOs o unidades documentales digitales. De igual forma que en una biblioteca se registran y se analizan (se catalogan, se clasifican, etc.) los libros que se adquieren y que engrosan el fondo, en una biblioteca digital de este tipo, se deben catalogar, clasificar, etc., es decir, se deben crear metadatos, para los documentos digitales que almacena en su sistema de información electrónico, ya que éstos son los documentos que realmente posee una biblioteca digital.

El procesamiento de la información que se realiza actualmente en la BVC se hace sobre el libro o documento impreso o en formato convencional. Esta circunstancia provoca el paradójico caso de que la colección descrita es "virtual" aunque tenga un formato físico tangible, en el sentido de que no está almacenada en la biblioteca. La descripción (catalogación) refiere a los documentos tangibles, donados o prestados para su digitalización, pero lo que se almacena es su trasunto digital, no el documento original⁴⁸.

⁴⁸ Esto puede producir situaciones tan paradójicas como que se sigan las normas de catalogación para cuestiones como la fuente principal de la catalogación de monografías, que constituyen una rémora innecesaria para el dinamismo que se supone en una biblioteca digital.

La crítica, más o menos explícita, que hemos hecho en esta tesis a la aplicación de técnicas tradicionales de catalogación y/o análisis documental sobre documentos digitales, se fundaba, sobre todo, en las descripciones excesivamente prolijas que requiere una catalogación MARC y que son, en muchos casos, inútiles en los documentos electrónicos. Sin embargo, lo que nos encontramos aquí no es una catalogación convencional de los documentos digitales que posee la BVC, sino una descripción del documento original en formato tradicional que en definitiva forma una colección paradójicamente "sí virtual" pero "no digital".

- Las búsquedas que permite la Biblioteca Virtual Cervantes en la actualidad, presentan esta realidad:
 - Las búsquedas en el *catálogo* sólo permiten localizar el documento completo según tres criterios (autor, título y materia) sobre una base de datos realizada en Access con referencia a campos MARC pero no en MARC. Esta circunstancia (no MARC) dificulta la conversión de la metainformación de la base de datos a metadatos propiamente dichos, ya que las *crosswalks* de MARC a distintos formatos de metainformación están perfectamente definidas⁴⁹ e implementadas como demuestra la tecnología aplicada a proyectos como CORC⁵⁰, e incluso, la conversión de MARC a XML⁵¹. Sin embargo en este caso, el mapeo de un formato a otro no tiene ninguna utilidad en el sentido de que los campos MARC en la base de datos bibliográfica de la BVC sólo aparecen de manera simbólica o referencial y no suponen en ningún caso una codificación ISO 2709, ni para la conversión de datos, ni para plantear la interoperabilidad

⁴⁹ Este tema ha sido tratado entre otros, por: Edward Gaynor. *From MARC to Markup... Op. cit.*, http://www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html; Judy Ahronheim, et al. *TEI/MARC "Best Practices"* [documento HTML]. Ann Arbor: Michigan University Library, 16 de junio de 2001. Disponible en: <http://www.lib.umich.edu/staff/ocu/teiguide.html> (consultado el 30 octubre de 2000).

⁵⁰ CORC lleva implícitos todos los desarrollos realizados en el seno del proyecto MANTIS de OCLC donde se incluye la modificación de registros MARC a DC pudiendo este último formato representarse también en RDF/XML. Keith Shafer. *Mantis, A Flexible Cataloging... Op. cit.*, <http://orc.rsch.oclc.org:6464/toolkit.html> Cfr. 4.1.

⁵¹ K.T. Lam hace un tratamiento completo de la representación de MARC en XML, de la conversión del formato bibliográfico al metalenguaje de descripción de información a través de XSLT, etc. K.T. Lam. *Moving from MARC to XML* [documento HTML]. Hong Kong University of Science and Technology, rev. 24 de mayo de 2001. Disponible en: <http://ihome.ust.hk/~lklkt/xml/marc2xml.html> (consultado el 19 de julio de 2001). Por otra parte, la Library of Congress ha desarrollado las DTDs de MARC en SGML y XML, una aplicación de XML para utilizar con datos MARC. Vid. <http://lcweb.loc.gov/marc/marcsgml.html>

entre la base de datos actual (si estuviese en MARC) y la base de metadatos⁵² (si existiese como tal).

- Las búsquedas en el *contenido de "libros"*, esto es, sobre el contenido el texto completo de los documentos que conforman la colección digital da la BVC, se realizan a partir del *search engine* HtDig y los resultados son poco precisos, ya que los criterios para el cálculo de la relevancia de HtDig están basados en HTML y el código HTML de los DLOs de la BVC se genera automáticamente. Aunque HtDig permite parametrizar las búsquedas a las metaetiquetas básicas de HTML (*keywords* y *description*) y asignarle pesos para la ordenación de resultados, el contenido de estas etiquetas en los códigos fuente HTML de los documentos de la BVC, son poco significativos para fundamentar las búsquedas⁵³:

```
<META NAME="Keywords" CONTENT="biblioteca, virtual,
Cervantes, literatura, gratis, El ingenioso hidalgo Don
Quijote de la Mancha, Cervantes Saavedra, Miguel de">
<META NAME="Description" CONTENT="Edición electrónica
incluida en la Biblioteca Virtual Miguel de Cervantes,
con un fondo de 3000 obras de literatura clásica
Española e Hispanoamericana, de acceso gratuito">
```

Nos hacen pensar que estas metaetiquetas están más dirigidas a los buscadores genéricos de la Red (SRII) que al *search engine* interno de la Biblioteca Virtual. Es evidente que el carácter genérico del contenido de las metaetiquetas está dirigido a la indización por motores de búsqueda de toda WWW. Esto se corrobora también porque todos los documentos HTML tiene en la cabecera de metadatos la etiqueta que permite a los robots que recorren la Red indizar el recurso pero no seguir los enlaces:

```
<META NAME="Robots" CONTENT="nofollow">.
```

⁵² *Vid. supr.* Nota 35.

⁵³ *Cfr.* Capítulo 4, nota 54.

- La recuperación de información es mejorable. La insuficiencia de la recuperación ya ha sido reconocida por la propia BVC que detecta la necesidad de realizar búsquedas más precisas y complejas para lo cual *se necesita un esquema consistente de marcado para toda la colección y técnicas de filtrado e indización para la búsqueda y recuperación de objetos textuales*⁵⁴, a lo que podemos añadir que se necesita un *schema* de metadatos y *schemes* consistentes de contenido para la recuperación de objetos de información (DLOs).
- En el *workflow* de la BVC que hemos descrito (9.2.2., Fig. 63) existe una mínima y escasa presencia de metadatos que no sirve para cimentar las búsquedas. No obstante la concreción y estabilidad del *workflow*, orientado a la producción masiva de DLOs, permite plantear la asignación de metadatos en el momento del marcado inicial del documento, que es el momento adecuado, en nuestra opinión, para crear la metainformación (ni antes de la existencia del DLO, es decir, sobre el libro o material original que se digitalizará, ni después, cuando ya se ha publicado en la Web). Es importante destacar el nivel de estructuración de la información a través del marcado descriptivo, así como el nivel de automatización de los procesos que, desde nuestro punto de vista son las fortalezas básicas en las que basar nuestro diseño de metadatos⁵⁵.
- La BVC presenta una casuística de DLOs muy interesante, tanto desde el punto de vista del formato (texto, voz, imagen, vídeo, etc.), como del contenido (obras literarias, tesis, homenajes, estudios críticos, etc.). Esta diversidad requiere un sistema de búsqueda más potente (que permita la búsqueda por tipo de

⁵⁴ Alejandro Bia, Andrés Pedreño. *Op. cit.*, p. 163.

⁵⁵ *Cfr. infr.* Tabla 17.

documento) como alternativa a la organización actual en secciones que requiere un alto conocimiento de la biblioteca por el usuario⁵⁶.

- El nivel de investigación, acreditada por la automatización de la producción de documentos digitales, de la Biblioteca Virtual *Miguel de Cervantes* puede trasladarse a la producción de registros de metadatos. Asimismo, el personal con el que cuenta esta biblioteca, entre bibliotecarios e informáticos, puede asumir el reto de la "reconversión" del sistema de organización de la información⁵⁷.

⁵⁶ Por ejemplo, tal y como está concebido en la actualidad el sistema de recuperación de información en la BVC, la única forma de buscar una imagen de Ángel González es acceder a la "biblioteca de imágenes" <http://www.cervantesvirtual.com/bib_imagenes/bibimagenes.shtml> y acceder a través de un listado de enlaces por autores a la entrada de "Ángel González", o bien saber previamente cual es el título exacto de, por ejemplo, una entrevista hecha a este autor y acceder a través del listado de títulos. Por otra parte, lo que contiene en realidad la "biblioteca de imágenes" es un conjunto de documentos multimedia (vídeos, etc.) no de imágenes fijas. También nos parece errónea la denominación de "listado de autores", porque el índice onomástico que aparece es de autores literarios no al autor de la imagen. Otra forma de llegar al fichero multimedia de Ángel González que estamos tomando como ejemplo, es buscar en el catálogo y seguir los enlaces hasta la ficha de autor donde aparece un *link* a dicho fichero multimedia; o bien acceder a la "biblioteca de autor" que tiene este personaje y a partir de ahí consultar todos los DLOs asociados a él.

Situaciones semejantes se producen al buscar otro tipo específico de DLO por ejemplo, un fichero de audio de o sobre Benedetti, o un tipo especial de información, por ejemplo un estudio crítico sobre Calderón. Los listados alfabéticos que suponen las distintas secciones de la biblioteca ("biblioteca de...") son ya demasiado extensos para un acceso ágil a la información y previsiblemente serán cada vez más amplios. Consideramos que la utilización de metadatos puede permitir un acceso más directo a los distintos DLOs y permitirá que sea el usuario el que seccione el catálogo según sus intereses, independizándole del *browsing* excesivo que supone la actual organización y de la necesidad de conocer la estructura de la biblioteca.

⁵⁷ Recordemos que de las ciento veintinueve personas que trabajan en la BVC un 14% son bibliotecarios e informáticos, esto supone un equipo de dieciocho personas que nos parece razonable para acometer la aplicación de metadatos sin que esto sea óbice para una dedicación mayor de personal. *Cfr. sup.* Nota 10.

- Por último, no existe en España ninguna biblioteca digital que base su organización en metadatos⁵⁸. En este sentido el proyecto de aplicación será pionero y servirá, dada la visibilidad, proyección y éxito de la BVC en la Web, de incentivo para los profesionales de la información españoles en la toma de conciencia de los estándares internacionales para la estructuración y recuperación de información en bibliotecas digitales.

9.3.2. Estado actual de la metainformación

Para evaluar el estado actual de los metadatos en la BVC, nos parece fundamental analizar la metainformación desde tres vertientes que serán a su vez, las mismas que presentemos en la propuesta de nuestro modelo:

- a) Los metadatos reales asociados al DLO XML-TEI incidiendo, a este respecto, en el momento de asignación de los mismos: el momento de creación y marcado del

⁵⁸ Existe un proyecto de investigación CICYT para crear un Sistema de Información Geográfica nacional basado en metadatos. Este proyecto surge de una comunidad virtual de usuarios (RedGeomática) albergada en RedIris, y se titula: *Desarrollo de servicios distribuidos de catálogo de información geográfica orientada a Internet y basada en estándares abiertos. Pasos efectivos hacia una Infraestructura Nacional de Información Geográfica*, y están implicadas las Universidades de Zaragoza, Jaume I, y Politécnica de Madrid. Vid. <http://redgeomatica.rediris.es/metadatos/index.htm> (consultado el 18 de julio de 2001).

Relacionado con los metadatos, aunque no expresamente en bibliotecas digitales, sino en sitios Web, podemos destacar el proyecto: *Recursos web i metadades* financiado por una ayuda económica de la *Divisió V per a projectes de recerca* de la Universidad de Barcelona y cuyo objetivo es comprobar las dificultades de aplicación del DC a la descripción de páginas de un sitio Web de la propia Universidad, tanto para la descripción de recursos como para su posterior búsqueda y recuperación. Sobre este proyecto ha aparecido recientemente (diciembre 2001) una publicación donde se describe: Assumpció Estivill Rius, et al. Recursos web i metadades: informe del projecte [documento HTML]. *bid: Biblioteconomia i Documentació*, nº 7, diciembre 2001. Disponible en: <http://www.ub.es/biblio/bid/07estiv1.htm> (consultado el 9 de diciembre de 2001).

También en el ámbito profesional, catalán, se ha presentado un informe sobre el estado de la cuestión de los metadatos, dirigido a los miembros del Consorci de Biblioteques Universitàries de Catalunya (CBUC) a fin de aportar los datos suficientes para la toma de decisiones sobre el uso de esta tecnología, así como proponer líneas de acción en este sentido dentro y fuera del consorcio. Lluïsa Núñez. *Metadates: informe per al CBUC* [documento PDF]. Barcelona: Biblioteca Digital de Catalunya, 2000. Disponible en: <http://www.cbuc.es/5digital/meta0007.pdf> (consultado el 22 de mayo de 2001).

DLO. A estos metadatos les vamos a denominar "metadatos recuperación-sistema" ya que, independientemente de cuándo se generen serán los que sirvan para la recuperación.

- b) Los metadatos asociados al documento editado en la WWW, esto es, la cabecera de metaetiquetas que aparece asociada al recurso en la Web y que es la que marcará, en su caso, la visibilidad de las obras en los buscadores. A este tipo de metadatos, teniendo en cuenta la distinción que hacemos en esta tesis⁵⁹, les denominaremos simplemente "metaetiquetas búsqueda-Web"
- c) Los metadatos que visualiza el usuario, es decir, información añadida, de la que dispondrán los usuarios de la BVC en la Web. A este tipo de metainformación la denominaremos "metadatos de usuario" o "metadatos-servicio", ya que se entienden como un servicio, el más importante, para el usuario.

Aunque de una u otra forma ya hemos hecho alusión⁶⁰ a todos estos tipos de metadatos⁶¹ que incluye la Biblioteca Virtual Cervantes, vamos a analizar ahora un ejemplo completo según estos tres puntos de vista. El DLO (o libro digital, en la denominación de la BVC) nº 001000 relativo a la obra de Francisco Navarro Villoslada, *El ante-cristo*⁶². Antes de analizar la metainformación es necesario precisar que, aunque esta obra, supone, en la edición digital de la BVC, 27 ficheros

⁵⁹ Cfr. Glosario, *metatag*.

⁶⁰ Vid. Ejemplo TEI Header (2) del apartado 5.2.2.1 como metadatos del sistema; Fig. 64. Ficha de catálogo de la BVC en la Web como metadatos de usuario o servicio, en ambos casos, relativos a la obra de *Las Esmeraldas* de Joaquín Dicenta y Capítulo 4, nota 54.

⁶¹ Aunque esta clasificación que apuntamos ahora no la consignamos al hablar de la tipología de metadatos (Cfr. 3.5.), sí hablamos sin embargo (3.4) de un enfoque funcional de los metadatos desde el punto de vista del sistema y desde el punto de vista del usuario. Este es el sentido que queremos dar ahora desde un punto de vista práctico: más que hablar tipos de metadatos, podemos hablar de niveles de utilización y visualización de los mismos.

⁶² Este documento se encuentra publicado en la BVC en: <http://www.cervantesvirtual.com/servlet/SirveObras/059596188953519652724313>

HTML (p0000001.htm...p0000026.htm, e index.htm), entendemos como DLO, en este caso, lo que en la BVC se llama libro digital, la obra completa ya que no se hace una descripción por distintas partes de los documentos y la obra completa es un solo fichero XML.

a) Metadatos recuperación-sistema

La metainformación del sistema, es decir, la información en la que se basa la recuperación en el catálogo de la biblioteca se extrae automáticamente, según explicamos (9.2.2., Fig. 63), de la base de datos en Access y se anexa, junto con algunos campos fijos, al documento sin marcado estructural. Esa información se traduce en el TEI Header, en el ejemplo que tomamos así:

```
<TEI.2>
<teiHeader day.created="21" month.created="7" year.created="1999">
  <fileDesc>
    <titleStmt>
      <title type="main">El ante-cristo</title>
      <author>Francisco Navarro Villoslada</author>
    </titleStmt>
    <publicationStmt>
      <publisher>Biblioteca Virtual Miguel de Cervantes
      Saavedra</publisher>
      <pubPlace>Universidad de Alicante</pubPlace>
      <idno>001000</idno>
      <availability status="free">
        <p>Copyright &copy; Universidad de Alicante, Banco
        Santander Central Hispano 1999-2001. Accesible desde
        http://cervantesvirtual.com</p>
      </availability><eDate day="21" month="7" year="1999"/>
    </publicationStmt>
    <sourceDesc> <p></p> </sourceDesc>
  </fileDesc>
  <profileDesc>
    <langUsage>
      <language id="es">Espa&ntilde;ol</language>
      <language id="la">Lat&iacute;n</language>
    </langUsage>
  </profileDesc>
  <revisionDesc>
    <change><eDate day="12" month="4" year="2001"/>
    <respStmt>
      <name>Sonia Jover S&aacute;nchez</name>
      <resp>Corrector</resp>
    </respStmt>
    <item>Etiquetado del texto en XML siguiendo las normas del
    TEI-Lite</item>
  </change>
  </revisionDesc>
</teiHeader>
```

En cuanto a los metadatos descriptivos, reflejados en el modelo TEI en el elemento <filedesc>, donde se debe describir tanto el propio documento electrónico como la fuente, sólo se describe muy brevemente el "libro electrónico" ya que el elemento sourcedesc aparece vacío sin ninguna información relativa a la fuente. Sólo se consignan dos de los tres elementos obligatorios: título y mención de responsabilidad (autor), la fecha de creación o publicación (availability) y el copyright, así como un elemento de número de identificación <idno>001000</idno> del documento que, como explicamos anteriormente, es el identificador del documento desde que se planifica su digitalización. Si bien es cierto que, a tenor de lo que refleja la DTD podrían incluirse otros datos como, *v. gr.*, la persona física o jurídica que dona o presta el libro para su digitalización, (elemento funder)⁶³.

Definición del elemento fileDesc en CERVANTE.DTD

```
<!ELEMENT fileDesc
  (titleStmt, publicationStmt, sourceDesc) >
<!ATTLIST fileDesc
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "fileDesc" >
<!ELEMENT sponsor
  (#PCDATA | ident | code | kw | abbr | address | date | name
  | num | rs | time | add | corr | del | orig | reg | sic
  | unclear | emph | foreign | gloss | hi | mentioned | soCalled
  | term | title | ptr | ref | xptr | xref | s | seg | formula |
  index | interp | interpGrp | lb | milestone | pb | gap |
  anchor)* >
<!ATTLIST sponsor
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "sponsor" >
```

⁶³ A continuación reflejamos la descripción de la cabecera TEI en la definición del Tipo de Documento CERVANTE.DTD, utilizada para marcar y codificar los documentos textuales en la BVC, en secciones significativas, comentando los aspectos más significativos a tenor de los metadatos TEIH que se consignan en el documento tomado como ejemplo (001000.xml) relativo a la obra de Francisco Navarro Villoslada, *El ante-cristo*. Para una mejor comprensión de las explicaciones que apuntamos aquí, *Vid.* 5.2.2.1 y Tabla 13 sobre la interpretación de los elementos principales de la TEIH en XML.

```

<!ELEMENT funder
  (#PCDATA | ident | code | kw | abbr | address | date | name
   | num | rs | time | add | corr | del | orig | reg | sic
   | unclear | emph | foreign | gloss | hi | mentioned | soCalled
   | term | title | ptr | ref | xptr | xref | s | seg | formula |
   index | interp | interpGrp | lb | milestone
   | pb | gap | anchor)* >
<!ATTLIST funder
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "funder" >
<!ELEMENT principal
  (#PCDATA | ident | code | kw | abbr | address | date | name
   | num | rs | time | add | corr | del | orig | reg | sic
   | unclear | emph | foreign | gloss | hi | mentioned | soCalled
   | term | title | ptr | ref | xptr | xref | s | seg | formula |
   index | interp | interpGrp | lb | milestone
   | pb | gap | anchor)* >
<!ATTLIST principal
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "principal" >
<!ELEMENT edition
  (#PCDATA | ident | code | kw | abbr | address | date | name
   | num | rs | time | add | corr | del | orig | reg | sic
   | unclear | emph | foreign | gloss | hi | mentioned | soCalled
   | term | title | ptr | ref | xptr | xref | s | seg | formula |
   index | interp | interpGrp | lb | milestone
   | pb | gap | anchor)* >
<!ATTLIST edition
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "edition" >
<!ELEMENT extent
  (#PCDATA | ident | code | kw | abbr | address | date | name
   | num | rs | time | add | corr | del | orig | reg | sic
   | unclear | emph | foreign | gloss | hi | mentioned | soCalled
   | term | title | ptr | ref | xptr | xref | s | seg | formula |
   index | interp | interpGrp | lb | milestone
   | pb | gap | anchor)* >
<!ATTLIST extent
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "extent" >
<!ELEMENT distributor
  (#PCDATA | ident | code | kw | abbr | address | date | name
   | num | rs | time | add | corr | del | orig | reg | sic
   | unclear | emph | foreign | gloss | hi | mentioned | soCalled
   | term | title | ptr | ref | xptr | xref | s | seg | formula |
   index | interp | interpGrp | lb | milestone
   | pb | gap | anchor)* >
<!ATTLIST distributor
  id ID #IMPLIED
  n CDATA #IMPLIED

```

```

    lang IDREF #IMPLIED
    rend CDATA #IMPLIED
    TEIform CDATA "distributor" >
<!ELEMENT authority
  (#PCDATA | ident | code | kw | abbr | address | date | name
  | num | rs | time | add | corr | del | orig | reg | sic
  | unclear | emph | foreign | gloss | hi | mentioned | soCalled
  | term | title | ptr | ref | xptr | xref | s | seg | formula |
  index | interp | interpGrp | lb | milestone
  | pb | gap | anchor)* >
<!ATTLIST authority
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "authority" >

```

La información sobre la codificación (<encodingDesc>) está excluida directamente en la reducción de la definición del tipo de documento XML-TEI que utiliza la BVC para marcar y codificar la información textual que edita electrónicamente (CERVANTE.DTD):

Definición del elemento `teiHeader` en CERVANTE.DTD

```

<!ELEMENT teiHeader
  (fileDesc, profileDesc, revisionDesc)64 >
<!ATTLIST teiHeader
  id ID #IMPLIED
  type CDATA "text"
  dtd-date (20010507) "20010507"
  day.created (1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12
  | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 |
  25 | 26 | 27 | 28 | 29 | 30 | 31) #REQUIRED
  month.created (1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
  12) #REQUIRED
  year.created (1998 | 1999 | 2000 | 2001 | 2002 | 2003 | 2004 |
  2005 | 2006 | 2007 | 2008 | 2009 | 2010) #REQUIRED
  status (new | update) "new"
  TEIform CDATA "teiHeader" >

```

Según esto podemos interpretar que no se le da importancia a la constancia de datos relativos a los niveles de codificación y análisis del texto, ni a la relación entre el fichero electrónico y la fuente en que se basa. Esto implica que *a priori*, no se

⁶⁴ Nótese cómo en la declaración de elementos que componen el TEIH en la DTD, no se incluye el elemento <encodingdesc>.

incluyen metadatos de los denominados administrativos por Gilliland-Swetland⁶⁵ o extrínsecos según Burnett, Bor Ng y Park⁶⁶. Sin embargo, sí están definidos en la DTD los elementos de la descripción de la codificación, tales como la finalidad por la cual se ha codificado el fichero electrónico (`projectDesc`), o cómo se ha tratado el fichero electrónico en cuanto a decisiones técnicas del proceso de edición digital (`editorialDecl`):

Definición del elemento `encodingDesc` en `CERVANTE.DTD`

```
<!ELEMENT encodingDesc
  (projectDesc*, samplingDecl*, editorialDecl*, tagsDecl?,
  refsDecl*, classDecl*, p*) >
<!ATTLIST encodingDesc
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "encodingDesc" >
<!ELEMENT projectDesc
  (p)+ >
<!ATTLIST projectDesc
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  default (YES | NO) "NO"
  TEIform CDATA "projectDesc" >
<!ELEMENT samplingDecl
  (p)+ >
<!ATTLIST samplingDecl
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  default (YES | NO) "NO"
  TEIform CDATA "samplingDecl" >
<!ELEMENT editorialDecl
  (p)+ >
<!ATTLIST editorialDecl
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  default (YES | NO) "NO"
  TEIform CDATA "editorialDecl" >
<!ELEMENT tagsDecl
```

⁶⁵ Anne J. Gilliland-Swetland. La definición de los metadatos. En: *Introducción a los metadatos: vías... Op. cit.*, p. 3.

⁶⁶ Katherine Burnett, Kwong Bor Ng, Soyeon Park. *Op. cit.*, p. 1214.

```

        (rendition*, tagUsage*) >
<!ATTLIST tagsDecl
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "tagsDecl" >
<!ELEMENT tagUsage
  (#PCDATA | ident | code | kw | abbr | address | date | name
  | num | rs | time | add | corr | del | orig | reg | sic
  | unclear | emph | foreign | gloss | hi | mentioned | soCalled
  | term | title | ptr | ref | xptr | xref | s | seg | formula |
  eg | bibl | biblFull | q | label | list
  | listBibl | note | figure | stage | table | text | index
  | interp | interpGrp | lb | milestone | pb | gap | anchor)* >
<!ATTLIST tagUsage
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  gi NMTOKEN #REQUIRED
  occurs NMTOKEN #IMPLIED
  ident NMTOKEN #IMPLIED
  render IDREF #IMPLIED
  TEIform CDATA "tagUsage" >
<!ELEMENT rendition
  (#PCDATA | ident | code | kw | abbr | address | date | name
  | num | rs | time | add | corr | del | orig | reg | sic
  | unclear | emph | foreign | gloss | hi | mentioned | soCalled
  | term | title | ptr | ref | xptr | xref | s | seg | formula |
  eg | bibl | biblFull | q | label | list
  | listBibl | note | figure | stage | table | text | index
  | interp | interpGrp | lb | milestone | pb | gap | anchor)* >
<!ATTLIST rendition
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "rendition" >
<!ELEMENT refsDecl
  (p)+ >
<!ATTLIST refsDecl
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  doctype NMTOKEN "TEI.2"
  TEIform CDATA "refsDecl" >
<!ELEMENT classDecl
  (taxonomy)+ >
<!ATTLIST classDecl
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "classDecl" >
<!ELEMENT taxonomy
  (category+ | ((bibl | biblFull), category*)) >
<!ATTLIST taxonomy
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED

```

```

        rend CDATA #IMPLIED
        TEIform CDATA "taxonomy" >
<!ELEMENT category
  (catDesc, category*) >
<!ATTLIST category
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "category" >

```

En la descripción del perfil, la única información no bibliográfica que se incluye es la relativa el idioma del texto (en el ejemplo del documento de Villoslada, latín y español), sin embargo como en el anterior elemento constitutivo de la cabecera TEI, podrían haberse explicitado otros elementos como, por ejemplo, clasificación del texto digital (`textClass`) y la codificación de la misma (`classCode`), o el uso de descriptores (`keywords`) en el caso de no utilizar ningún lenguaje controlado:

Definición del elemento `profileDesc` en CERVANTE.DTD:

```

<!ELEMENT profileDesc
  (langUsage) >
<!ATTLIST profileDesc
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "profileDesc" >
<!ELEMENT creation
  (#PCDATA | ident | code | kw | abbr | address | date | name
  | num | rs | time | add | corr | del | orig | reg | sic
  | unclear | emph | foreign | gloss | hi | mentioned | soCalled
  | term | title | ptr | ref | xptr | xref | s | seg | formula |
  index | interp | interpGrp | lb | milestone
  | pb | gap | anchor)* >
<!ATTLIST creation
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "creation" >
<!ELEMENT langUsage
  (language)+ >
<!ATTLIST langUsage
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  default (YES | NO) "NO"
  TEIform CDATA "langUsage" >
<!ELEMENT language
  (#PCDATA)* >
<!ATTLIST language

```

```

n CDATA #IMPLIED
lang IDREF #IMPLIED
rend CDATA #IMPLIED
id (ar | ca | de | el | en | es | eu | fr | gl | it | iw | la
| oc | pt | ro | ru) #REQUIRED
wsd ENTITY #IMPLIED
usage NMTOKEN #IMPLIED
TEiform CDATA "language" >
<!ELEMENT textClass
(classCode | catRef | keywords)* >
<!ATTLIST textClass
id ID #IMPLIED
n CDATA #IMPLIED
lang IDREF #IMPLIED
rend CDATA #IMPLIED
default (YES | NO) "NO"
TEiform CDATA "textClass" >
<!ELEMENT keywords
(term+ | list) >
<!ATTLIST keywords
id ID #IMPLIED
n CDATA #IMPLIED
lang IDREF #IMPLIED
rend CDATA #IMPLIED
scheme IDREF #IMPLIED
TEiform CDATA "keywords" >
<!ELEMENT classCode
(#PCDATA | ident | code | kw | abbr | address | date | name
| num | rs | time | add | corr | del | orig | reg | sic
| unclear | emph | foreign | gloss | hi | mentioned | soCalled
| term | title | ptr | ref | xptr | xref | s | seg | formula |
index | interp | interpGrp | lb | milestone
| pb | gap | anchor)* >
<!ATTLIST classCode
id ID #IMPLIED
n CDATA #IMPLIED
lang IDREF #IMPLIED
rend CDATA #IMPLIED
scheme IDREF #IMPLIED
TEiform CDATA "classCode" >
<!ELEMENT catRef
EMPTY >
<!ATTLIST catRef
id ID #IMPLIED
n CDATA #IMPLIED
lang IDREF #IMPLIED
rend CDATA #IMPLIED
target IDREFS #REQUIRED
scheme IDREF #IMPLIED
TEiform CDATA "catRef" >

```

Llama especialmente la atención que no conste más metainformación en la cabecera del documento marcado de acuerdo a la DTD, pues la fuente de extracción de estos datos es la información catalográfica de la base de datos en Access, completada por los bibliotecarios sobre el documento original que se va a digitalizar, donde sí consta la clasificación (CDU) y la asignación de materias, etc. propias de la

catalogación convencional. La justificación evidente a que no consten esos datos de materias la encontramos en que, a pesar de que añadir a los parámetros de extracción de metainformación estos datos no conllevaría ninguna complicación, no tendrían ninguna utilidad, pues no existe un *search engine* en la BVC que trabaje con estos metadatos directamente y tampoco son visibles para los motores de búsqueda de la WWW. No obstante sí se añaden metadatos de contenidos fijo⁶⁷ en el momento de creación del documento HTML con la finalidad de que sean analizados por los SRII genéricos de Internet.

Por último, en la cabecera de los documentos marcados TEI, según se muestra en el ejemplo del documento de Villoslada, se incluyen en el elemento relativo a la descripción de la revisión, detalles relacionados con la actualización y corrección del texto (fecha y la persona encargada de la revisión, así como su responsabilidad y la fecha en que se realizó) conforme a lo que se expresa en la DTD.

Definición del elemento `revisiónDesc` en `CERVANTE.DTD`

```
<!ELEMENT revisiónDesc
  (change+) >
<!ATTLIST revisiónDesc
  id ID #IMPLIED
  n CDATA #IMPLIED
  lang IDREF #IMPLIED
  rend CDATA #IMPLIED
  TEIform CDATA "revisiónDesc" >
```

Como decíamos al hablar de TEI en el capítulo 5, si una TEIH crea de tal forma que contenga el mismo contenido que un registro MARC, con la misma adherencia a la práctica catalográfica, requerirá el mismo nivel de destreza que para la catalogación bibliotecaria. Pero en este caso, esa destreza bibliotecaria se aplica previamente para describir y catalogar los documentos en formato tradicional, con lo

⁶⁷ Como los que señalamos anteriormente. P. ej:

```
<META NAME="Description" CONTENT="Edición electrónica incluida
en la Biblioteca Virtual Miguel de Cervantes, con un fondo de
3000 obras de literatura clásica Española e Hispanoamericana,
de acceso gratuito">
```

Cfr. infr. b) Metaetiquetas de búsqueda-Web, asociadas a la obra *El ante-cristo* que estamos tratando.

cual, una extracción de los metadatos descriptivos y del perfil sería sencilla. Sin embargo el modelo de creación de metadatos sobre el documento original demuestra que, si bien se pueden crear metadatos relativos al contenido intelectual de las obras, así como el autor, título etc. e información sobre la fuente, el hecho de que se trabaje sobre el documento tradicional imposibilita la asignación de información intrínseca y administrativa del documento digital (esto es, el DLO que realmente posee la BVC, y que podemos considerar que forma parte de su colección).

En cualquier caso, los metadatos que hemos denominado de recuperación-sistema, partiendo de la teoría general de los metadatos, deben ser los más importantes en este tipo de bibliotecas digitales por varias razones:

- Porque en ellos descansará el sistema de recuperación local. Supuestamente el motor de búsqueda (*search engine*) local indizará los elementos que aparezcan en esas cabeceras.
- Porque servirán para extraer los otros dos tipos de metainformación: la que visualizará el usuario para anticipar el acceso al documento digital y las metaetiquetas para la visibilidad en los motores de búsqueda.

b) Metaetiquetas búsqueda-Web

Las etiquetas <META> que contienen todos los documentos de la BVC tal y como se visualizan en la WWW son (siguiendo el ejemplo del DLO textual de la obra digitalizada, marcada y editada digitalmente *El ante-cristo*):

```
<HEAD>
<!-- Biblioteca Virtual Miguel de Cervantes Saavedra -->
<!-- Obra: El ante-cristo -->
<!-- Autor: Navarro Villoslada, Francisco -->
<!-- Nro.Catalogo: 001000 -->
<TITLE>El ante-cristo - Biblioteca Virtual Miguel de
Cervantes</TITLE>
<META NAME="Title" CONTENT="El ante-cristo - Navarro
Villoslada, Francisco">
<META NAME="Author" CONTENT="Biblioteca Virtual Miguel de
Cervantes">
<META NAME="Owner" CONTENT="webmaster@cervantesvirtual.com">
```

```
<META NAME="Revisit" CONTENT="60 days">
<META NAME="Keywords" CONTENT="biblioteca, virtual, Cervantes,
literatura, gratis, El ante-cristo, Navarro Villoslada,
Francisco">
<META NAME="Description" CONTENT="Edición electrónica incluida
en la Biblioteca Virtual Miguel de Cervantes, con un fondo de
3000 obras de literatura clásica Española e Hispanoamericana,
de acceso gratuito">
<META NAME="Robots" CONTENT="nofollow">
<META NAME="Language" CONTENT="Spanish">
<link rel="SHORTCUT ICON" href="/favicon.ico">
```

La información bibliográfica que aparece en la cabecera del documento HTML, una vez publicado en la Web, está incluida claramente a través de comentarios (<!-->) pues en las etiquetas <META> propiamente dichas (author, title, description, y keywords) no aparece la información que realmente describe la obra. Concretamente, el contenido que tiene la propiedad description, es idéntico para todos los documentos de la BVC⁶⁸. Además siempre se incluye el valor "nofollow" en la propiedad "Robots" porque, si bien no existe en la BVC ninguna política de restricción para que los robots indiquen cada página, sí se establece que no recorran todos los enlaces jerárquicamente, lo que supondría que, una vez detectada una página se indizaría toda la biblioteca.

Si analizamos la utilidad de estas metaetiquetas para aumentar la visibilidad en sistemas de búsqueda genéricos de la WWW, los resultados, a tenor de lo que explicamos en el capítulo 7 sobre la aprehensión de las etiquetas <META> por ellos, son previsibles, ya que no las tienen en cuenta para el cálculo de la relevancia, e incluso en algunos casos se obvian por completo en la indización. Así, por ejemplo si realizamos la búsqueda (adaptada convenientemente a cada lenguaje de interrogación) "el ante-cristo" AND Villoslada, por ejemplo, en Altavista <<http://www.altavista.com>> obtenemos, como único resultado, la traducción al español de las normas TEI, justamente realizada por Manuel Sánchez Quero, uno de los

⁶⁸ Por poner un ejemplo, la obra de Lope de Vega de título fonéticamente igual aunque de grafía diferente, *El antecristo*, también tiene como valor de la etiqueta description: *edición electrónica incluida en la Biblioteca Virtual Miguel de Cervantes, con un fondo de 3000 obras de literatura clásica Española e Hispanoamericana, de acceso gratuito*, que como en el resto de las obras, no refleja el contenido de las mismas, sino la misión de la BVC.

investigadores del equipo de informática de la Biblioteca Virtual Cervantes⁶⁹ en que se menciona un ejemplo con esta obra. Si realizamos la misma búsqueda en Google <<http://www.google.com>>, además ese resultado aparece otro sorprendente: una página de las bibliotecas rurales argentinas <<http://www.biblioteca.org.ar/titulo.asp?texto=e>> en la que, a través de *scripts* de Java dan acceso a las obras de la biblioteca virtual Cervantes, pero no en la propia BVC.

El ejemplo demuestra que la utilidad de las metaetiquetas para la Web general es muy escasa y concretamente, tal y como están en la BVC, creadas automáticamente, etc., aún menos. Aunque los motores de búsqueda las tuviesen en cuenta fielmente para la indización y la relevancia de sus resultados, la información que reflejan éstas en la BVC no es descriptiva del documento digital, ni tampoco lo es en sus elementos bibliográficos más básicos (autor, título, materias-keywords).

c) Metadatos de usuario-servicio

La información que recibe el usuario, a través de la "ficha de la obra"⁷⁰ es de dos tipos o pertenece a dos ámbitos informativos dentro de la proyección de la BVC como servicio.

- Por un lado, información bibliográfica mínima:

Autor/a: Navarro Villoslada, Francisco (1818-1895)
Título: El ante-cristo

⁶⁹ Lou Burnard, C. M. Sperberg-McQueen. *TEI Lite: Una introducción al Text Encoding for Interchange*, Junio de 1995. [documento HTML]. Traducción de Manuel Sánchez Quero. Alicante: Universidad de Alicante, rev. 5 de noviembre de 2001. Disponible en: http://www.alu.ua.es/m/msq1/Trad_TEI-Lite.htm (consultado el 7 de diciembre de 2001).

⁷⁰ La ficha relativa a la obra que estamos tratando se encuentra en: <http://www.cervantesvirtual.com/FichaObra.html?Ref=1000> Se trata de un documento HTML igualmente, y como tal, como ya hemos comentado (*Cfr.* Capítulo 4, nota 54) tiene asociada sus propias metaetiquetas en el código fuente de la misma (metainformación tipo: *metaetiquetas búsqueda-Web*) generada, como en el resto de los documentos, automáticamente en el momento de su publicación en la Web. La fuente de datos para crear automáticamente estos documentos HTML (ficha) es base de datos bibliográfica realizada en Access y no la cabecera TEI mínima asociada al documento.

Nota: Edición digital a partir de Obras Completas, Iruña (Pamplona), Mintzoa, 1990, t. V, pp. 25-292.

Materias:

CDU: (82-311.6) Histórica.

Encabezamiento de materia: Novela histórica española - Siglo 19º

CDU: 821.134.2-311.6"18"

Los metadatos descriptivos son muy básicos: autor⁷¹, título y nota de la fuente y las páginas en las que está incluido en el documento impreso. Por otra parte los metadatos dependientes del contenido (clasificación e indización) están controlados, sin embargo no son campos buscables en la interfaz de consulta ni sustentan ninguna forma de recuperación por materias. La búsqueda por materias queda relegada a la navegación a partir del catálogo de materias; donde encontrar la obra de Villoslada, nos supondría una navegación de seis enlaces (con la correspondiente carga de las seis páginas), y buscar posteriormente entre una lista de veinte novelas históricas que tiene clasificadas en este género la BVC⁷².

- Por otra parte, se consignan también datos relativos a la biblioteca como portal o servicio donde se proporciona metainformación al usuario, en este caso, datos añadidos a la frecuencia de consulta de la obra y sobre la posibilidad de interactuar con otros usuarios a través de los foros. La información de este tipo que aparece en la ficha-web del libro de Villoslada es exactamente, la siguiente:

Esta obra ha sido consultada en 291 ocasiones.

Foro de debate de la obra:

Acceso al foro (0 participantes).

Añadir un comentario al foro.

Acceso a direcciones de interés recomendadas por nuestros visitantes (0 participantes).

Recomendar una dirección de interés (URL) relacionada con la obra.

⁷¹ Nótese el control de autoridades en los encabezamientos de autor y a través de ellos, la vinculación del proceso técnico de la BVC al mundo bibliotecario tradicional.

⁷² Esto en el supuesto de que supiéramos que *El ante-cristo* es una novela histórica y de que conociéramos mínimamente la CDU en la que se basa el catálogo de materias. Si el catálogo de CDU ha sido poco o nada consultado en las bibliotecas tradicionales, salvo por los propios bibliotecarios, menos aún se va a utilizar para navegar en el entorno más dinámico e intuitivo que impone la Web.

Los resultados de la recuperación de información fruto de esta metainformación, son inadecuados o al menos, insuficientes. Para llegar a esta conclusión debemos argumentar dos niveles de RI: 1) a nivel local, es decir, la evaluación de los sistemas de búsqueda dentro de la propia BVC y 2) a nivel global, esto es, el la recuperación genérica de la WWW.

1) En el sistema de **búsqueda local** podemos distinguir asimismo, dos subniveles:

- La búsqueda local cualificada, donde interrogación directa en el formulario de consulta del catálogo (base de datos bibliográfica de la BVC) es poco ágil pues requiere, desde nuestro punto de vista, una navegación ulterior excesiva, por las fichas de autores, títulos o materias, teniendo que seleccionar, en muchos casos, la obra deseada de largas listas de enlaces con un único criterio alfabético de títulos o autores.

En cuanto a la navegación directa (sin interrogación previa) a partir de los diferentes catálogos⁷³, supone también un largo y tedioso proceso de *browsing* donde la navegación dirigida, se abandona, igualmente, en listados alfabéticos heterogéneos.

- En cuanto a la **búsqueda local en texto completo**, es también insuficiente, pues el volumen de información es muy grande y tiene un contenido literario muy alto, de tal forma que la búsqueda a través de htDig es útil en muy pocos casos.

2) La **búsqueda global en Internet** también podemos analizar dos subniveles, en consonancia con nuestra tesis:

- La **búsqueda global en SRII**. Aunque este tipo de búsqueda es difícil de valorar, pues los sistemas de recuperación y su funcionamiento trascienden a

⁷³ Catálogos alfabéticos (de títulos o autores) y CDU (materias): <http://www.cervantesvirtual.com/catalogo.shtml> Vid. Fig. 65.

la BVC, si podemos evaluar la metainformación consignada en sus documentos a tenor del funcionamiento genérico de los motores de búsqueda. Las metaetiquetas embebidas automáticamente en el código fuente de los documentos no coadyuvan de ninguna manera a la visibilidad de la colección de la BVC, aunque sí de la biblioteca virtual en sí misma⁷⁴.

- La **búsqueda global integrada**. La búsqueda integrada se refiere a la posibilidad de recuperar documentos de la BVC entre sistemas heterogéneos, bien con otras bibliotecas digitales semejantes (como por ejemplo el Proyecto Gutenberg), bien con bibliotecas convencionales basadas en MARC dando lugar a un servicio de búsqueda híbrido (como por ejemplo el catálogo de la propia biblioteca de la Universidad de Alicante que trabaja con Unicorn). En este sentido, el estado actual de la metainformación en la Biblioteca Virtual *Miguel de Cervantes*, no permite integrar búsquedas fundadas en la interoperabilidad, pues su sistema de "metainformación" no es compatible con ningún estándar (p. ej., MARC, DC, RDF) ni con ningún protocolo (p. ej., Z39.50) que permita basar la búsqueda en la interoperabilidad.

A raíz de este análisis sobre el estado actual de los metadatos y las posibilidades que estos brindan a distintos niveles de recuperación, nuestra propuesta es la aplicación de metadatos que potencien las posibilidades de gestión y recuperación de DLOs de la colección de la BVC de una forma ágil a todos los niveles que señalamos.

⁷⁴ Incluso se produce el caso, como demostramos en el ejemplo de búsqueda en Google y Altavista, consignado anteriormente, ficticias, virtuales en el sentido estricto de la palabra porque no poseen la colección que referencian, como el caso de las bibliotecas rurales argentinas que hemos mencionado, tienen más visibilidad, en cuanto a los documentos individuales, que la propia biblioteca Cervantes.

9.3.3. Valoración y selección de un modelo de metadatos

Como vimos en la segunda parte de esta tesis doctoral⁷⁵, el despliegue de modelos, esquemas y tentativas de normalización de metadatos hace que exista una miscelánea de posibilidades para adoptar un formato de metadatos que se adapte a las necesidades de una biblioteca digital en cuestión, en nuestro caso, a la Biblioteca Virtual Cervantes. Para ello, partimos del principio de normalización y de aceptación. Esto es, trataremos de proponer un modelo basado en alguno de los estándares existentes, aceptado por la comunidad bibliotecaria, antes que desarrollar un nuevo formato *ad hoc* para dicha biblioteca.

En un proyecto de biblioteca digital particular, una vez que los objetivos del acceso están definidos, se suelen elegir los estándares de metadatos que mejor se ajusten a los objetivos y técnicas adoptadas. Según este principio, algunos modelos pueden parecer evidentes, verbigracia, en el caso de la BVC, asumir que si la información está codificada en TEI-XML, los metadatos deben estar siguiendo el estándar de la TEIH. Sin embargo es preciso considerar también las posibilidades de procesamiento que tienen dichas cabeceras así como el hecho que ya hemos apuntado, de que la BVC no sólo alberga documentación textual codificada TEI, sino una gran diversidad de DLOs no-textuales.

Antes de proponer la estructura de metainformación más adecuada, es preciso tener en cuenta una serie de criterios para evaluar los modelos existentes⁷⁶, partiendo de que la complejidad y variedad informativa de la BVC requiere prever las

⁷⁵ Vid. Capítulos 5 y 6.

⁷⁶ Los criterios que hemos tenido en cuenta están basados en:

- Jean Hudgins, Grace Agnew, Elizabeth Brown. *Op. cit.*, p.45-47.
- DESIRE. *Information Gateways...*, *Op. cit.*, <http://www.desire.org/handbook/print4.html> [sección 2.3. *Metadata formats*, p. 34-43. en la versión impresa o en el URL <http://www.desire.org/handbook/2-3.html>].
- Ya-ning Chen, et al. A Case Study in Designing Chinese Metadata. *Online Information Review*, vol. 24, nº 3, 2000, p. 232.

consecuencias de la recuperación de información digital dentro de la estructuración de diferentes procesos de digitalización y de diferentes productos o DLOs. Así, el modelo de metainformación tiene que:

- Recoger los requisitos técnicos básicos y característicos de los procesos de digitalización (parámetros y sistemas de digitalización) ya que en este caso la biblioteca es productora/editora de su colección y estos procesos conllevan metainformación administrativa y técnica interesante para la gestión del subsistema editorial.
- Debe de ser un modelo flexible que permita presentar diversas propuestas de edición y explotación de documentos electrónicos (sistemas de almacenamiento, sistemas de servicio de imágenes, etc.).
- Además se tienen que tener en cuenta cuestiones básicas como:
 - El contexto de aplicación, esto es, la temática de la información que alberga nuestra biblioteca digital. En este caso, multitemática, con un marcado sesgo académico y humanístico.
 - Las características del grupo de usuarios al que va dirigido: un usuario "universal".
 - El material que compone la colección y el/los formato/s de la misma: material totalmente digital o digitalizado (texto, audio, imagen y multimedia, además de la compleja unidad informativa de los denominados "portales temáticos" y/o "de autor", que contiene tanto documentos textuales relativos a las obras, textos HTML contextuales y otros DLOs relacionados).
 - El carácter amigable del modelo, teniendo en cuenta tanto a los diseñadores y gestores del proyecto, como al usuario final, en este sentido apuntamos a semánticas de propósito general y sencillas como el Dublin Core.

- El proceso de entrada de los datos así como el *workflow* de producción para establecer las condiciones de representación y almacenamiento. Aquí es importante pensar en varias cuestiones: quién va a crear los metadatos, en qué momento de los flujos de trabajo del subsistema editorial, qué sintaxis se adoptará para expresar los metadatos, qué *scheme/s* se utilizarán para valores controlados, qué herramientas de software se utilizarán para ultimar todo el proceso de creación-gestión y recuperación de metadatos, dónde y cómo se almacenarán los metadatos y cómo se accederá a ellos.
- Y sobre todo, se tiene que evaluar rentabilidad del proyecto partiendo de la relación inversión (tiempo y dinero)-mejora previsible de las posibilidades del servicio de recuperación⁷⁷.

Haciendo una valoración amplia y genérica de todas estas cuestiones, así como todo lo que hemos apuntado a lo largo de la tesis, consideramos que la solución idónea es utilizar, la semántica del Dublin Core en una sintaxis RDF, por las razones que indicamos a continuación:

- La semántica del Dublin Core nos parece la más adecuada por su carácter general, por su nivel de normalización (ANSI/NISO Z39.85) además de su proximidad al mundo bibliotecario y de sus posibilidades de adaptación a las necesidades de una colección a través de perfiles de aplicación (*application profile*) mezclando distintos *namespaces*⁷⁸.
- RDF es la infraestructura que posibilita una codificación de intercambio y reutilización de estructuras de metadatos y además facilita una visión completa de los datos a través de mecanismo que, como los *namespaces*, soportan

⁷⁷ Obviamente nosotros no evaluamos este aspecto en este trabajo. Sin embargo pensamos que es necesario establecer varios muestreos y opciones antes de la aplicación que proponemos aquí ahora.

⁷⁸ Vid. Glosario, *application profile*, *namespace*.

convenciones comunes de semántica, sintaxis y estructura. Además está basado en XML y su integración en una estructura TEI-XML es posible. Por otra parte, cada vez más, la tendencia es a crear colecciones heterogéneas, bien virtuales, bien reales digitalizadas almacenadas en local, que requieren, en el caso de la BVC, asociar metadatos por ejemplo a la colección de documentos en PDF a través de XAP (*eXtensible Metadata Platform*⁷⁹), o elementos del esquema DIG35 para imágenes.

RDF es una PAS del W3C y su nivel de aceptación y adopción es creciente para la organización de bibliotecas digitales. Esta circunstancia hace previsible la interoperabilidad futura con otras bibliotecas digitales o con catálogos convencionales en MARC. Además la posibilidad de crear RDFS para el vocabulario, permite proyectar la integración de metadatos asociados a la materia añadidos, en un futuro al sistema de recuperación de información local.

- La creación de perfiles de aplicación (*application profiles*) es la última propuesta en el diseño en metadatos, apoyada por la DCMI⁸⁰ y por SCHEMAS en el ámbito europeo. Sirven para definir esquemas de metadatos consistentes en elementos elegidos de uno o más *namespaces*, combinados en un mismo registro y optimizados para una aplicación local. Las características fundamentales que nos ayudan a proponer un perfil de aplicación son: que pueden representar uno o más espacios de nombre, pueden refinar elementos de un estándar de metadatos adaptándolos a las necesidades concretas, complejas y particulares de la BVC, permiten también diseñar *schemes* para los valores controlados que es, la

⁷⁹ XMP es la iniciativa de la empresa Adobe para incluir metadatos en el fichero binario de los documentos pdf. Más información: <http://www.adobe.com/products/xmp/main.html> Ponemos este ejemplo, ante la imposibilidad de consignar toda la casuística de formatos (imagen, multimedia, texto, etc.) de la BVC y la cantidad de esquemas que debemos combinar en el diseño definitivo del perfil de aplicación, ya que los documentos PDF tienen la una circunstancia especial de producción que no conlleva marcado descriptivo, y es por otra parte una realidad frecuente en la BVC, en algunos casos, de manera exclusiva como en portal de la BN de Chile.

⁸⁰ Cfr. 5.2.1.2.

tendencia, de claro cariz tradicional bibliotecario que impera en la organización y clasificación de sus DLOS. Además, iniciativas de interoperabilidad entre bibliotecas y sistemas de información digital (como el OAI) demandan perfiles de aplicación que combinen elementos de distintos esquemas de metadatos normalizados al mismo tiempo que aseguran la interoperabilidad y satisfacen requisitos especiales para una colección. Este planteamiento es, en principio, el más idóneo para la BVC en el sentido de que la interoperabilidad no sólo será necesaria en un futuro de proyección de integración de sus registros con catálogos MARC o con otros catálogos de proyectos de bibliotecas digitales, sino que la interoperabilidad es un imperativo apriorístico para elegir un sistema local de recuperación de información común y optimizado para la complejidad informativa de la colección de la BVC.

- La codificación del DC en XML/RDF es cada vez más estable y previsiblemente será una recomendación en enero de 2002⁸¹.
- Además los perfiles de aplicación pueden basarse tanto en esquemas (*schemas*) RDF como en *schemas* puros de XML⁸². Sin embargo la opción del esquema RDF, desde el punto de vista normativo, está avalada también por el proyecto europeo SCHEMAS.

⁸¹ Dave Beckett, Eric Miller. Expressing Simple Dublin Core in RDF/XML [documento XHTML]. DCMI, 28 de noviembre de 2001. Disponible en: <http://www.dublincore.org/documents/2001/11/28/dcmes-xml> (consultado el 6 de diciembre de 2001).

⁸² Lagonze y Hunter presentan una aproximación que combina el conocimiento semántico de los esquemas RDF con las restricciones estructurales explícitas de los esquemas XML, que demuestran la complementariedad de ambos. Si bien se podría diseñar un esquema RDF específico para la BVC consideramos que es necesario esperar a que se convierta en una recomendación (recordemos que es recomendación candidata desde marzo del 2000). *Vid.* Jane Hunter, Carl Lagonze. Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles [documento HTML]. Brisbane: University of Queensland, Distributed Systems Technologies, rev. 22 de abril de 2001. Disponible en: <http://archive.dstc.edu.au/RDU/staff/jane-hunter/www10/paper.html> (consultado el 3 de agosto de 2001).

- Por último, la investigación en transformaciones a través de XSLT en la BVC, nos permite pensar que se puede automatizar el proceso de creación visualización de metadatos, al igual que se hace con la producción de libros.

Con este planteamiento tan específico sería difícil contar con una aplicación comercial que contemplase todos estos requisitos, salvo en el caso de herramientas integrales como MetaStar de BlueAngel o DigiTooLibrary de Ex-Libris⁸³, que en cualquier caso nos obligarían a plegar nuestros intereses a sus funcionalidades. No obstante, dado el carácter de investigación que envuelve el proyecto de la BVC desde sus orígenes, consideramos que la adopción de soluciones comerciales no estaría de acuerdo con la política del proyecto, además desconocemos los precios y la rentabilidad exacta de dichas aplicaciones. Por ello y sin que obste la creación de aplicaciones *ad hoc* en consonancia con nuestras observaciones, se propone adoptar el conjunto de herramientas *Extensible Open RDF* (EOR⁸⁴), un prototipo de aplicaciones desarrolladas y distribuidas siguiendo el modelo de fuente abierta (*open source*) que soportan la construcción y gestión de una base de datos para agregar, interrogar, gestionar y visualizar metadatos en RDF. Entre otras ventajas EOR está constituido de aplicaciones como: Apache 1.3.14, Jakarta-Tomcat Server/3.2.1 (JSP 1.1; Servlet 2.2;), Blackdown J2SDK 1.3.0, MySQL 3.23.33⁸⁵, MM.MySQL JDBC Driver, la revisión de la interfaz de programación de aplicación de Stanford⁸⁶, Xalan-Java.

EOR está diseñado para validar RDF, construir y buscar almacenes de datos RDF (bases de metadatos), convertir datos RDF, etc. Además facilita un lenguaje de

⁸³ Vid. 4.3.2.3. y también 4.3.3.

⁸⁴ Vid. The EOR Toolkit: <http://eor.dublincore.org>

⁸⁵ La concepción de RDF en triples, permite hacer un paralelismo con SQL y la representación de bases de datos (tabla, columna, fila) con RDF (clase, propiedad, recurso).

⁸⁶ Recordemos la investigación técnica en el desarrollo de aplicaciones llevado a cabo por la Universidad de Standford bajo los auspicios del proyecto DLI.

interrogación sencillo (lógica booleana, unión e intersección, etc.). Creemos que puede ser un punto de partida adecuado para sustentar los tres niveles de metainformación deseables:

a) Metadatos recuperación-sistema

En cuanto a los metadatos del sistema se deben crear a partir del documento digital, no del documento-fuente y almacenarse en la base de datos EOR, así como la posibilidad de embeber también el código RDF del recurso bajo la propia estructura TEI de los documentos. Se pueden crear a través de una plantilla en Java (con un funcionamiento semejante, p. ej. a Reggie⁸⁷) que facilite su asignación a los bibliotecarios y que se integre con el sistema de almacenamiento.

Todo ello permitirá crear un modelo de búsqueda basado en campos DC pudiendo realizar búsquedas más amplias, cualificadas por campos tan necesarios en la BVC, como p. ej., el tipo de documento. En cualquier caso, el código RDF asociados al documento de Villoslada, utilizando sólo el esquema DC para la codificación semántica de etiquetas, sería:

```
<?xml version = "1.0"?>
<RDF xmlns = "http://www.w3.org/TR/1999/REC-rdf-syntax-
19990222#"
  xmlns:DC = "http://metadata.net/dstc/DC-10-ES/#">
  <Description xml:lang="es"
  about="http://www.cervantesvirtual.com/Servlet/SirveObras/05959
6188953519652724313">
  <DC:Title.Main>El ante-cristo</DC:Title.Main>
  <DC:Creator.PersonalName>Francisco Navarro Villoslada
  </DC:Creator.PersonalName>
  <DC:Subject>Novela historica española, siglo XIX</DC:Subject>
  <DC:Description>Versión digital de la obra de Francisco Navarro
  Villoslada, El ante-cristo, realizada por la Biblitoeca Virtual
  Miguel de Cervantes. Novela histórica que refleja el debate
  religioso, social y político de la España del siglo XIX.
  </DC:Description>
  <DC:Publisher.CorporateName> Biblioteca Virtual Miguel de
  Cervantes
  </DC:Publisher.CorporateName>
  <DC:Publisher>Banco Santander Central Hispano</DC:Publisher>
  <DC>Date.Available DC:Scheme="ISO8601">1999-07-21
  </DC>Date.Available>
```

⁸⁷ Vid. 4.3.2.1 y Fig. 25.

```
<DC:Type>texto</DC:Type>
<DC:Format>html</DC:Format>
<DC:Identifier DC:Scheme="URI">
http://www.cervantesvirtual.com/Servlet/SirveObras/059596188953
519652724313 </DC:Identifier>
<DC:Source>Obras completas. Iruña (Pamplona), Mintzoa, 1990, t.
5, p. 25-292</DC:Source>
<DC:Language>
  <Bag>
    <li DC:Scheme="RFC1766">es</li>
    <li DC:Scheme="RFC1766">la</li>
  </Bag>
</DC:Language>
<DC:Relation.IsPartOf>Obras completas. Iruña (Pamplona): 1990
</DC:Relation.IsPartOf>
<DC:Coverage.PeriodName>1845</DC:Coverage.PeriodName>
<DC:Coverage.PlaceName>España</DC:Coverage.PlaceName>
<DC:Rights>http://www.cervantesvirtual.com</DC:Rights>
</Description>
</RDF>
```

b) Metadatos búsqueda-Web

En cuanto a los metadatos de búsqueda Web, planteamos dos opciones alternativas:

- La conversión a metadatos DC en HTML en el momento de su publicación. De igual forma que el documento marcado TEI-XML se publica en la Web en HTML a través de *parsers XSLT* y genera unas metaetiquetas, como hemos visto en el apartado anterior, poco útiles, se puede generar el código RDF abreviado embebido en HTML. De esta forma, en un futuro puede ser visible a motores de búsqueda de la WWW que soporten los nuevos planteamientos de la Web Semántica.
- Insertar los metadatos a través de la opción de embeberlos *on the fly* en el momento de la visualización del documento por ejemplo a través de *scripts* en Perl⁸⁸ que encuentren el registro apropiado de metadatos en la base de datos.

⁸⁸ Esta técnica se la propuesta por el diseño de metainformación de las páginas del propio proyecto eLib del que hemos hablado en varios momentos de esta investigación. Cfr. <http://www.ukoln.ac.uk/services/elib/metadata.html>

Además de para la búsqueda global, esos metadatos pueden tenerse en cuenta para mejorar los resultados de la búsqueda local sobre el texto completo si se cambia el indizador htDig por Swish-E⁸⁹ (también *open source*) que reconoce la semántica básica del DC. Esta es una forma transitoria de cualificar también las búsquedas en texto completo vinculándolas igualmente a la existencia de metadatos, a la par que se puede prever un aumento de la visibilidad global en la Web.

c) Metadatos de usuario

En cuanto a los metadatos de usuario, consideramos que la gestión y codificación de metadatos permitirá abandonar el complejo e inapropiado (por su vinculación a la biblioteca tradicional) sistema de fichas de obra y de autor. Sin embargo, la experiencia en la producción automática de las fichas HTML puede generar visualizaciones de metadatos con un nivel de detalle más amplio que la información de las fichas catalográficas, en la página Web (una sencilla y posible visualización es la que proponemos en la Fig. 67). Además, a través de estilos XSL y de *parsers* XSLT, se pueden lograr distintas presentaciones⁹⁰ según niveles de usuario.

⁸⁹ Swish-E: <http://swish-e.org> Cfr. Eric Lease Morgan. Comparing Open Source Indexers [documento HTML]. *Infomotions Musings*, May 29, 2001. Disponible en: <http://www.infomotions.com/musings/opensource-indexers> (consultado el 20 de junio de 2001).

⁹⁰ Un caso semejante a la visualización diversa de información bibliográfica a través de XSL lo describe Robert Ferrer. University of Illinois the Federation of Digital Libraries: Interoperability Among Heterogeneous Information Systems. En: *Digital Libraries: Philosophies... Op. cit.*, p. 102-103.

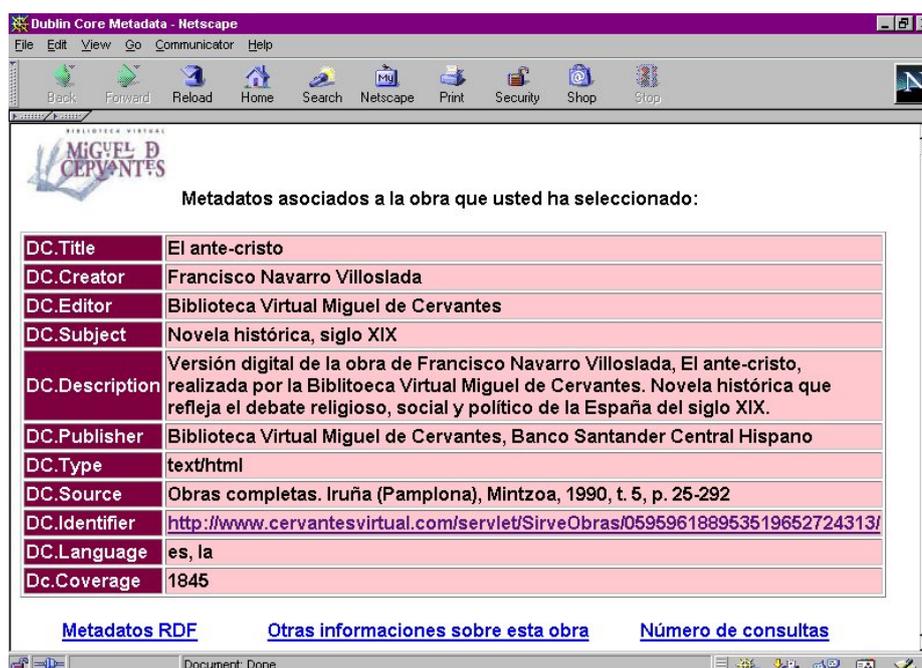


Fig. 67. Propuesta de visualización de metadatos de usuario en la BVC

A continuación planteamos el rediseño de los flujos de trabajo adaptándolos a la asignación eficaz de metadatos que hemos planteado. El cambio fundamental propuesto es la vinculación del proceso de asignación de metadatos al proceso de marcado y producción del documento digital, alejando éste del libro impreso o documento fuente ya que, como dijimos, una biblioteca digital de las características de la Cervantes debe describir y crear metadatos para la colección que realmente posee. La base/s de metadatos EOR, serán el núcleo de la generación de todos los niveles de metainformación consignados en la BVC, a partir de ella se extraerá: tanto la cabecera RDF que se puede embeber en el documento digitalizado sin marcar o con un marcado mínimo automático, como para los metadatos de usuario y los destinados a la búsqueda en texto completo (tanto local, como por SRII).

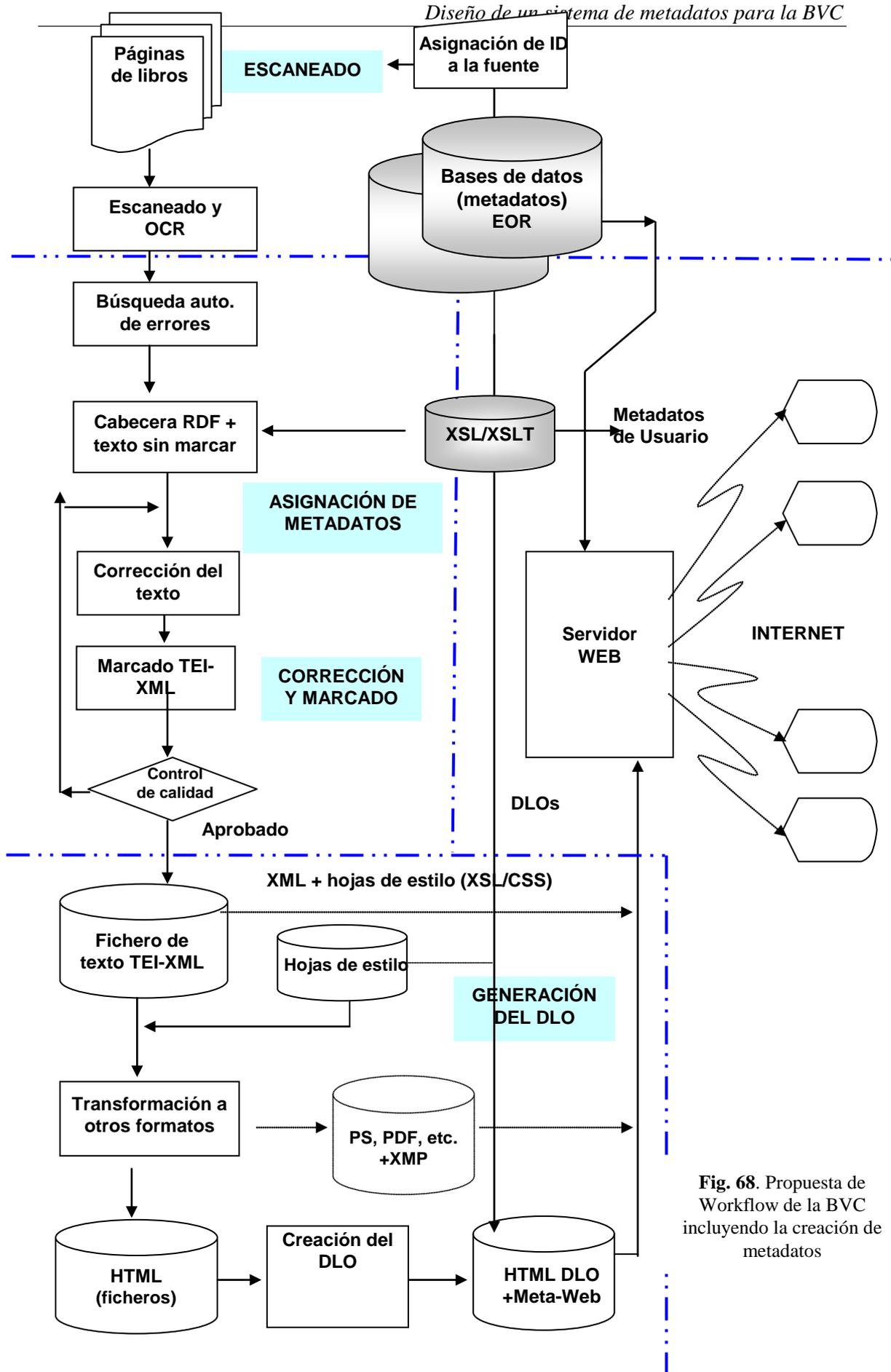


Fig. 68. Propuesta de Workflow de la BVC incluyendo la creación de metadatos

9.3.4. Prospectiva y viabilidad del proyecto

Hasta ahora hemos apuntado algunas propuestas de diseño de un sistema de información basado en metadatos para la Biblioteca Virtual *Miguel de Cervantes* que contempla, en su composición como sistema y como concepto en la nueva Sociedad de la Información, un subsistema de producción o edición digital. Sin embargo no hemos hecho planteamientos definitivos porque serán parte de un desarrollo posterior consensuado por el equipo informático y bibliotecario de la misma, pero sí es importante plantear una prospectiva de aplicación de metainformación en general, que incluya éstas y otras reflexiones.

Las decisiones de embeber metadatos en un fichero individual o almacenarlos en una base de datos centralizada pueden no ser excluyentes como apuntábamos, pero dependerá de la dificultad real de integración en el sistema de la biblioteca virtual. Los beneficios de mejorar el acceso a la información deben balancearse no obstante, con relación al esfuerzo, y por ello se deben replantear a tenor de los presupuestos o tentativas de financiación. En muchos casos, según demuestran las publicaciones que relatan procesos de aplicación de metadatos a bibliotecas digitales, la solución de más bajo coste inicialmente, puede ser la de incluir los metadatos con el marcado y asignarlos en el momento de producción del DLO, sin embargo, esta opción es beneficiosa en el caso de que sean sitios sencillos y pequeños, sin embargo para un proyecto de la magnitud de la BVC puede ser inviable, máxime cuando ya existen más de 7000 documentos producidos y gestionados en el sistema actual con datos mínimos de metadatos TEIH que no se utilizan para la recuperación.

En el caso de la BVC, a pesar de que el momento idóneo de asignación de metadatos es la creación del documento digital⁹¹, sí creemos que el almacenamiento debe de ser externo pues el fondo está formado no sólo por documentos textuales, e

⁹¹ Normalmente, como se ha defendido en esta tesis, y así consta como opinión consensuada en la bibliografía sobre el tema, los metadatos deben responder a una ECIP (*Electronic Cataloging In Publication*).

incluso, algunos documentos textuales son ficheros binarios, como por ejemplo PDF o PS. También se puede contemplar la opción de crear vínculos a los metadatos, tal y como veíamos en el ejemplo de la página Web del DCMI⁹² con el elemento LINK, de forma externa, de tal forma que los metadatos RDF⁹³ constituyan un almacenamiento independiente y buscable en el sistema distribuido o federado.

La mayor dificultad de aplicar un esquema RDF (RDFS) para constitución y definición de perfiles *ad hoc* no es todavía recomendación del W3C, y como muestra la experiencia y hemos reflejado en esta tesis, su aplicación en el mercado no se regularizará hasta que no se convierta en PAS. Por ello las aplicaciones como EOR para la creación de bases de datos RDF que fundamenten la recuperación basada en metadatos o Protegé⁹⁴ para la definición de esquemas, etc. se encuentran en fase de desarrollo. La Web Semántica es todavía un proyecto y la explotación de las ventajas de RDF para la estructuración de datos y metadatos y es sólo operativa en contextos cerrados, finitos de información. Empero es una inversión de futuro codificar así los metadatos para la interoperabilidad con otras bibliotecas digitales, que basen sus sistemas de organización de la información en RDF/XML. Mientras tanto, la propuesta que hacemos permitirá anotar el contenido Web visible en la Red a partir de metaetiquetas HTML que pueden ser el prolegómeno de la Web Semántica⁹⁵.

Finalmente, para analizar la viabilidad de la aplicación de metadatos, hemos hecho un resumen DAFO (Debilidades, Amenazas, Fortalezas y Debilidades) de la Biblioteca Virtual *Miguel de Cervantes* sobre esta propuesta (Tabla 17) que se irá

⁹² Cfr. Capítulo 4, nota 87.

⁹³ Vid. Fig. 67, también consigna esta opción de visualizar metadatos en RDF para usuarios especializados.

⁹⁴ Cfr. Capítulo 5, nota 220 y Fig. 54.

⁹⁵ Uche Ogbuji. An introduction to RDF: Exploring the Standard for Web-based Metadata [documento HTML]. *IBM Web Architecture Articles*, December 2000. Disponible en: <http://www-106.ibm.com/developerworks/Webservices/library/w-rdf/index.html> (consultado el 8 de enero de 2001).

reelaborando en el futuro a tenor de experiencias piloto o de avances en la concepción del diseño que planteamos en este capítulo.

FORTALEZAS	DEBILIDADES
<ul style="list-style-type: none"> - Concepción generalizada de la edición de documentos a partir del marcado descriptivo XML-TEI. - Experiencia en la automatización de procesos de producción y visualización de datos y documentos, trasladable a la generación de metadatos. - Existencia de un subsistema de planificación tecnológica sólido. - Interés por la investigación y la innovación. - Recursos humanos y materiales disponibles. - Calidad general del proyecto. 	<ul style="list-style-type: none"> - Excesiva vinculación a la biblioteca tradicional. - Excesiva vinculación al documento fuente. - No-concepción del documento como objeto de información (DLO). - Escasa estructuración de la poca metainformación asignada a los documentos (campos basados en MARC pero no MARC) que hace impracticable la conversión. - Catalogación convencional. - Dependencia excesiva del <i>browsing</i> para la recuperación.
OPORTUNIDADES	RIESGOS
<ul style="list-style-type: none"> - Interés reconocido por el personal informático de la BVC - Estabilización creciente de la sintaxis RDF. - Estabilidad consolidada de la semántica DCMES (ANSI/NISO Z39.85) - Previsible que el esquema RDF se convierta pronto en recomendación. - Consideración de RDF en el diseño de nuevos protocolos y en la redefinición de protocolos tradicionales como Z39.50. - Gran desarrollo de sistemas <i>Open source</i>. 	<ul style="list-style-type: none"> - La aplicación de metadatos suspenda temporalmente el funcionamiento actual. - No-aceptación del modelo por los bibliotecarios. - Convivencia temporal de dos sistemas de recuperación. - Dificultad de implementación en el sistema de la metainformación relativa a los servicios (metadatos de usuario).

Tabla 17. Resumen DAFO de la BVC ante la aplicación de metadatos

A pesar de las debilidades y los riesgos apuntados, el desarrollo tecnológico general, el componente humano y la responsabilidad en innovación que confluyen en la Biblioteca Virtual *Miguel de Cervantes*, hacen viable este proyecto de aplicación de metadatos. Consideramos que merece la pena probar este diseño, realizando muestreos sobre secciones de la colección aprovechando, la división en “portales” temáticos de la misma, que permitirán generar un sistema piloto de implementación a un conjunto limitado de DLOs.

Las expectativas de investigación, en torno a los metadatos, como muestra esta tesis son interminables, pudiendo ampliarse no sólo al recurso sino también al contenido. En un futuro se puede plantear también la creación de una ontología en RDF que permita generar un sistema de metadatos y de recuperación basados en materias. RDF es un formato flexible y adecuado para la integración con el modelo de biblioteca digital local pero tiene una proyección global y es el primer estadio para constituir la interoperabilidad con catálogos preexistentes en MARC o con otros catálogos de metadatos. La BVC amerita el esfuerzo de convertirse en pionera del cambio de paradigma en la gestión de contenidos enteramente digitales.

CAPÍTULO 10:

CONCLUSIONES

Como vaticinaba Licklider, en su obra *Libraries of the Future*, alrededor del año 2000 la información y el conocimiento son tan importantes como era la movilidad a mediados de los 60. Lo que no podía predecir este autor hace treinta y cinco años era que la información iba a caracterizar la nueva sociedad del siglo XXI, ni que el pilar tecnológico de esa sociedad iba a ser Internet.

Estamos en la segunda fase del sueño de la Web que Tim Berners Lee relata en su obra *Tejiendo la Web*. La primera parte de su sueño ya se ha cumplido y la Web se ha convertido en el medio más potente de colaboración entre las personas, donde todo el mundo tiene, al menos potencialmente, un acceso inmediato e intuitivo a la información, no sólo para navegar sino también para crear. En la segunda parte del sueño, las máquinas se vuelven capaces de analizar los datos que hay en la Web. El camino hacia la Web Semántica que debería hacer eso posible, no ha hecho más que empezar. Esa Web que puedan entender las máquinas llegará gracias a la implantación de una serie de avances técnicos y acuerdos sociales, gracias al desarrollo de la informática, de agentes inteligentes y, por qué no también, de los avances en la normalización de la gestión de la información. Por ello, en esta tesis hemos dado cuenta del papel que tienen las bibliotecas y los bibliotecarios en esa nueva organización y estructuración de la información electrónica. El lenguaje común que permitió recrear la primera parte del sueño de Berners Lee fue HTML, que facilitó y sigue facilitando que los ordenadores representen y compartan hipertexto. El lenguaje que permitirá a los ordenadores compartir y representar datos e información semántica (metadatos), entendibles por máquina y por el hombre, es RDF. En este trabajo reflejamos la importancia de la creación y gestión de metadatos

en el primer estadio de la Web Semántica: la aplicación de esa metainformación en contextos finitos de DLOs, más concretamente, en las denominadas bibliotecas digitales o virtuales.

Tal como decíamos en el capítulo 2, a raíz de la hipótesis que exponía Blair adaptando un aserto de Hegel, la cantidad de información accesible en el nuevo e inabarcable "sistema de información" de Internet, traerá, y está trayendo ya, cambios cualitativos en lo relativo a la organización y acceso a la información. Esta hipótesis genérica la hemos corroborado a lo largo de este trabajo. En este apartado recogeremos las conclusiones a las que hemos llegado a partir de nuestras hipótesis iniciales, consignando de forma sumaria las contribuciones que hemos ido apuntando en esta tesis doctoral, así como las expectativas de futuro de este ámbito de la investigación. De todo ello, destacamos o colegimos lo siguiente:

1. La estructuración de la información a través del establecimiento y creación de sistemas de metadatos, lejos de ser una mera catalogación (*vino viejo en odres nuevos* —como diría Gradman—, o *una vieja respuesta a un viejo problema* —en palabras de Lucas—), constituye la mayor promesa para controlar este universo de información heterogéneo y caótico que supone Internet, además de ser la oportunidad para redefinir el papel del profesional de la información en el contexto de los paradigmas de información emergentes suscitados por la Red, como las bibliotecas digitales o virtuales.

Toda catalogación implica metadatos, pero no todos los metadatos implican catalogación. Esta afirmación es la única que puede aclarar el debate científico entre la catalogación y la asignación de metadatos. Desde nuestro punto de vista, la creación de metadatos difiere de la catalogación, fundamentalmente en el objeto (DLOs en la primera, y documentos "tradicionales" en la segunda). Los metadatos deben ser a la biblioteca digital —con todas las salvedades que hemos señalado a lo largo de esta investigación tanto en lo relativo a la cobertura del concepto, como a la legibilidad humana y por máquina que implican, como al cambio de paradigma en los procesos documentales— lo que el análisis documental a la biblioteca tradicional.

2. La palabra que mejor define el estado actual de los metadatos es *variedad*:
 - Variedad de posturas conceptuales: El concepto de metadatos es aún ambiguo y necesitará todo un proceso de afianzamiento y consenso por parte de los distintos sectores de investigación implicados.
 - Variedad de aplicaciones y usos. Aunque aquí hemos destacado su utilidad para la recuperación de información en la Red y más concretamente en las bibliotecas digitales, son muchas las utilidades que se le reconocen a la metainformación.
 - Variedad de tipos y variedad de criterios para establecer esas tipologías.
 - Variedad de formas de creación, almacenamiento y gestión de esos metadatos.
 - Variedad de aplicaciones y herramientas para editar y gestionar la metainformación. En algunos casos, simplemente para aumentar la visibilidad de las publicaciones electrónicas, y en otros, para sustentar verdaderas bibliotecas digitales.
 - Variedad de modelos, formatos, esquemas y/o estándares, diferentes y heterogéneos tanto en la sintaxis, semántica y contenido, como en su nivel de alcance y dominio de aplicación.
 - Variedad de arquitecturas para la recuperación de información en Internet basada en esas estructuras de representación de la información, denominadas metadatos.
3. En el desarrollo de los metadatos, como tecnología y teoría subsidiaria del mercado de documentos electrónicos para la recuperación de información en bibliotecas digitales, se advierten dos fases: una primera etapa protagonizada por la proliferación de propuestas de formatos con pretensión de estándares *de facto* para optimizar la recuperación de la información en toda la Red (1995-1998), en la que se utilizan también modelos de metadatos preexistentes como IAFA o

SOIF; y una segunda etapa que persigue la formalización de estándares de metadatos de reconocimiento internacional y constitución consorciada, y que tiende a la interoperabilidad en el escenario del acceso global a la información, entre distintos sistemas y/o bibliotecas digitales (1999-2001).

4. La explosión de formatos, modelos y estándares de metadatos, así como todo el trabajo surgido en torno a ellos nos ha servido para comprobar que la organización de la información, la "catalogación" en sentido amplio, que en muchos ámbitos se presenta como una actividad arcana e ininteligible que pertenece sólo al mundo de los bibliotecarios y archiveros, se está tomando como un punto de inflexión por una comunidad científica mucho más amplia. Los problemas que plantea la recuperación de información, han trascendido los muros estrictos de las bibliotecas y los centros de documentación ante el nuevo desafío de la información electrónica, y preocupan tanto a informáticos, como a filósofos, educadores, conservadores del patrimonio, topógrafos o geólogos, así como a un sin fin de investigadores que aspiran a organizar y optimizar sus sistemas de información.
5. Teniendo en cuenta que los recursos de Internet experimentan un crecimiento exponencial, su descripción por profesionales de la información o catalogadores, sería ilusoria, impracticable y, desde el punto de vista del coste, prohibitiva. Por otra parte, la extracción de metadatos de forma totalmente automatizada es —en la actualidad, contando con herramientas como Klarity— inexacta o imprecisa para un procesamiento coherente y descriptivo de la información electrónica.
6. Las soluciones a los problemas que se suscitan en torno a la creación de metadatos provienen de dos vertientes distintas: de un lado, la recuperación genérica de información en Internet (motores de búsqueda etc.) donde el propio creador del documento, a través de esquemas sencillos y de herramientas o interfaces fáciles de usar, pueda crear metadatos adecuados para la descripción y el establecimiento del contenido de los recursos y de las relaciones semánticas entre éstos; y de otro lado, los sistemas de información electrónica, sean bibliotecas digitales, *subject gateways*, u otros contextos finitos de información,

donde sí es preciso contar con personal especializado, o una nueva figura de "catalogador-versado-en-tecnologías", que pueda aprovechar la potencialidad de esquemas de metainformación más complejos, que permitan desarrollar verdaderos servicios organizados de información con valor añadido.

7. El hecho de que se inste a la formulación de propuestas o a la formalización de información (RFP/RFI) sobre los sistemas y aplicaciones de metadatos denota que, a pesar del reconocimiento de la importancia del modelo de gestión de información que posibilitan los metadatos en el contexto de la Sociedad de la Información digital, aún existe un gran desconocimiento del mercado del software para la creación de sistemas basados en la información sobre la información.
8. El software para la creación y gestión de metainformación es, en nuestra opinión, abundante pero, desde el rigor y utilidad que se le presume en esta tesis a los metadatos para la organización de la información electrónica, aún insuficiente. La mayoría de las herramientas para la edición de metadatos (sean *templates*, o software cliente) están volcadas a la "promoción de la Web", más que a la creación de sistemas descriptivos sólidos que coadyuven a la búsqueda de información de calidad. Las aplicaciones existentes son un reflejo de varias de las hipótesis que manejamos en esta tesis:
 - La "moda" de los metadatos.
 - La falta de estándares suficientemente sólidos y la preeminencia, desarrollo, estabilidad y popularidad del Dublin Core como norma semántica de metainformación.
 - Las soluciones parciales a medida, fruto de proyectos de investigación o de bibliotecas digitales particulares.
 - El estado aún incipiente de muchos de los temas relacionados con los metadatos.
 - Sobre todo, la necesidad de diferenciar entre el uso de metainformación para incrementar la relevancia de la búsqueda global en Internet

(metaetiquetas), frente al uso de metainformación para constituir sistemas sólidos de calidad o verdaderas bibliotecas digitales (metadatos).

9. Los sistemas habituales de recuperación de información en Internet, destinados a la recuperación genérica de documentos Web, se basan en la indización automática del texto completo, realizada por robots que recorren sistemáticamente la Red. La recuperación en estos sistemas es deficiente desde el punto de vista de que no contemplan ningún aspecto semántico, lo que se traduce en respuestas cuantitativamente inabarcables e irrelevantes desde el punto de vista del concepto de *relevancia subjetiva* del usuario. Salvo algunas excepciones (*v.gr.* Oingo u Open Directory) estos sistemas no se basan en metadatos para la formación de sus bases de datos y para la recuperación de información; a lo sumo, aunque cada vez menos, tienen en cuenta algunas metaetiquetas de la cabecera HTML.
10. El almacenamiento de los metadatos para la recuperación sigue siendo un problema difícil de resolver. Se debe avanzar en la creación de sistemas que puedan pasar de la extracción automática de metainformación codificada con el propio documento, a una base de metadatos que facilite la recuperación.
11. Existen en la actualidad múltiples modelos y estándares de metadatos aplicables a un tipo o formato de documento o a un área temática concreta. Aunque todos están adoptando, o han adoptado ya, una sintaxis XML, en cuanto a la semántica, existe un claro predominio del esquema de la DCMI, bien por la adopción del conjunto de elementos del DC, bien por el influjo que tiene este modelo en el desarrollo de otros de propósito específico basados en él. Sin embargo, el uso eficaz de metadatos requiere convenciones o normalizaciones relativas a la semántica, sintaxis y estructura de la metainformación.
12. La investigación en la aplicación y normalización de metadatos, puede aportar un valor añadido a los sistemas de recuperación de información en la Red, sin olvidarnos del papel que desempeñan las tecnologías *per se*, de cuya evolución y de la asimilación de estos modelos *documentales* puede depender el éxito o el fracaso de sistemas de recuperación basados en metadatos.

13. Los sistemas de información distribuida en Internet y las bibliotecas digitales deben fundamentarse en metadatos normalizados abocados a la interoperabilidad, más que en un único estándar formal de metadatos. Esto no obsta para que se desarrollen normas nacionales e internacionales al respecto que contribuirán, en el momento que sean asumidas por los desarrolladores de software, a mejorar la representación y recuperación del conocimiento.
14. A pesar de que existen esquemas e incluso normas formales para la descripción de recursos de Internet, su aplicabilidad exige persuadir a los creadores de información Web para que utilicen dichos esquemas, así como a los desarrolladores de aplicaciones para que asimilen tales normas en sus productos.
15. Los formatos e iniciativas que hemos tratado en este trabajo, así como el resto de modelos surgidos en el seno de dominios de información concretos, suponen un avance para enfrentarse al desafío de una recuperación de información relevante y eficaz en sistemas de información distribuidos como la WWW. Sin embargo, y como defendemos en esta tesis, la aplicación de los metadatos, hoy es factible y realmente operativa en contextos finitos, abarcables de información, sea su finitud física o temática, y alcanzan su mayor utilidad en el contexto de las bibliotecas digitales.

La creciente verticalización de la información en Internet, la especialización temática entendida desde distintos puntos de vista (comercio electrónico, comunidades virtuales, etc.) corrobora esta tesis de aplicabilidad de metadatos en contextos de información determinados *a priori*.

16. Las bibliotecas digitales son un contexto finito y operativo de recursos de información electrónicos *organizados*. La aplicación de un sistema de metadatos sirve para organizar los DLOs que constituyen sus colecciones, contribuyendo a mejorar la recuperación de información dentro de ellas, así como a potenciar, en un futuro, el intercambio de datos fuera de los límites definidos de la biblioteca o colección digital.

Los metadatos son el sustento de sistemas informativos digitales de calidad y el elemento que permite que grandes colecciones de documentos y datos electrónicos funcionen como bibliotecas organizadas.

17. Si la biblioteca tradicional es un conjunto *organizado* de documentos, la biblioteca digital es un conjunto *organizado* de documentos entendidos como objetos de información (DLOs), que constituyen un nuevo sistema de información en el marco de la comunicación social global. Así, distinguimos dos tipos fundamentales de biblioteca digital:

- Las bibliotecas digitales editoras y poseedoras de su colección digital. La formación de colecciones se basa en documentos impresos que se seleccionan, se digitalizan, se estructuran a través del marcado y se organizan a través de metadatos creados antes de la edición digital del documento. La recuperación se basará en los metadatos extraídos automáticamente, entre otros modelos, y formando bases de datos que se podrán interrogar en función de los metadatos que contienen asignados *a priori*.
- Bibliotecas digitales temáticas (*subject gateways*) basadas en el concepto de "colección virtual" distribuida y en la función de acceso, en vez de custodia, que rompen las barreras físicas de localización y de propiedad de la información difundida en las bibliotecas convencionales. Estas bibliotecas seleccionan DLOs a los que adscriben metadatos para una representación post-publicación de la información digital temática que albergan. La recuperación se basará en la interrogación de esos metadatos asignados *a posteriori*.

18. En cualquier caso, la conformación de bibliotecas digitales debe basarse —como ocurre en las bibliotecas tradicionales— en la selección y/o en la organización. La recuperación de la información puede fundarse, por un lado, en las estructuras lógicas de metadatos basados en el recurso, y por otro, en un control del vocabulario y en la formación de ontologías del ámbito temático que representan metadatos basados en materias.

19. El proverbial problema de la recuperación por materias, es más grave en Internet, al no existir la posibilidad de que se construya un sistema de representación del conocimiento multilingüe y multitemático que abarque todo el conocimiento distribuido de la Red. Sin embargo sí es factible fundamentar la recuperación por materias en sistemas de organización del conocimiento (clasificaciones, tesauros, etc.) en el contexto de bibliotecas digitales, normalmente temáticas.
20. La globalización y la asunción de los retos de la Sociedad de la Información a nivel internacional provocará (está provocando ya) que el número y la diversidad de recursos informativos en la World Wide Web continúe creciendo de manera exponencial. Paralelamente los proyectos de codificación numérica o digitalización del conocimiento continuarán aumentando y el concepto de biblioteca digital, como conjunto organizado de recursos electrónicos, será cada vez más una necesidad y una realidad para dar acceso a ese conocimiento.
21. Sólo resta esperar al siguiente estadio en la madurez en la organización y la estructuración de la información para convertir la Web en una Web Semántica. De momento, la metainformación muestra su utilidad para la recuperación de información de calidad en contextos finitos, abarcables de documentos digitales.
22. Los metadatos, en el marco de la Sociedad de la Información, son un elemento normativo más, de naturaleza técnica, que acompaña y sustenta su desarrollo en el entramado político y filosófico de la democratización del acceso a la información a través de la tecnología; protagonizarán además, el cambio de paradigma en la gestión de la información.
23. La inestabilidad, la falta de consenso sobre el nivel de detalle de descripción y la complejidad de los metadatos, nos hacen pensar que la solución del problema de la recuperación global y consistente dista mucho de ser una realidad a corto plazo. Para que los metadatos sirvan de soporte de una recuperación global en la Red, es preciso que se conviertan en estándares *de facto* y se desarrollen *crosswalks* para la interoperabilidad entre distintas bibliotecas digitales, y entre bibliotecas digitales y bibliotecas tradicionales.

24. La investigación en torno a la interoperabilidad terminológica para mejorar la recuperación por materias se centra en el desarrollo del RDF/XML para representar estructuras jerárquicas y compartir esquemas de organización del conocimiento, que darán lugar a un complejo sistema semántico de recuperación de información basada en metadatos orientados a materias.

25. De forma temática y en general, el ámbito de información más desarrollado, en cuanto al establecimiento y aplicación de metadatos a verdaderas colecciones o bibliotecas digitales, es la información geoespacial. Esa madurez se puede ver en:

- La solidez normativa del CSDGM, aprobado por el *Federal Geographic Data Committee* (FGDC) desde 1994 (versión 1) y norma ISO desde el año 2000, o la norma europea ENV 12657:1998 (*Geographic Information - Data description – Metadata*).
- La variedad, nivel de desarrollo e internacionalización de las aplicaciones y herramientas existentes para crear, validar y gestionar este tipo de metadatos geoespaciales.
- La evolución de los sistemas específicos para la organización del conocimiento en Red (ontologías, tesauros, etc.) sobre esta materia, que dan coherencia semántica a los metadatos.
- La constitución efectiva de múltiples bibliotecas digitales (*clearinghouses*) a nivel internacional sobre información geográfica y espacial, que basan su establecimiento y difusión informativa en un modelo de metadatos para la recuperación de información.

26. Si utilizamos el criterio geográfico para evaluar la aplicación de estructuras de metadatos para la recuperación de información, podemos destacar fundamentalmente los siguientes países o focos geográficos:

- Australia, tanto desde el punto de vista de la adaptación y creación de estándares propios de metadatos (EdNA, ANZLIC, AGLS), como desde la propia adopción comercial de estos esquemas, siendo predominantes las empresas australianas que desarrollan software específico (Metabrowser,

Metamanager, Klarity, MetaStar). Es preciso subrayar la existencia de una política nacional clara de gestión de metadatos, liderada por la *National Library of Australia* (NLA) y asumida por todos los servicios de información temática (*subject gateways*) de este país. Destaca también una investigación sólida en este tema, protagonizada sobre todo, por la labor del RDU.

- Estados Unidos, a través de diferentes iniciativas como las de la RLG, el GILS o la propia *Library of Congress* u OCLC, que se han preocupado por desarrollar y mantener estándares de metadatos. Asimismo, destaca en este país la promoción y la continuidad en las investigaciones sobre bibliotecas digitales dentro de la DLI (*Digital Library Initiative*).
- En Europa ameritan mención especial en este sentido: el Reino Unido, sobre todo a través del trabajo e investigación de la UKOLN y del apoyo del *Joint Information System Committee* (JISC) en la aplicación de las TIC a los sistemas de información; así como los Países Escandinavos, por el carácter pionero en la asunción de los metadatos para la gestión de información electrónica a través del *Nordic Metadata Project* (1996-1998).
- En cuanto a nuestro país, son muy escasas y tímidas aún las experiencias de aplicación de metadatos y el único proyecto destacable está vinculado, como cabría esperar, a la información geoespacial. Detectamos además la carencia de una investigación profunda en el tema y la ausencia total de actividades coordinadas.
- A nivel supranacional, son relevantes las iniciativas de la Unión Europea, que inició su labor con relación a los metadatos a través del proyecto *Telematics for Libraries* (1990-1998) durante la vigencia del IV Programa Marco de I+D, celebrando tres talleres monográficos y financiando diversos proyectos como NEDLIB o BIBLINK. Actualmente el seguimiento de la investigación y aplicación de metainformación se está realizando a través del IST del V Programa Marco (1998-2002), en el contexto de su tercera acción clave, *contenidos y herramientas multimedia*, a través de proyectos como TEL (*The European Library*), Delos, Cultivate o METAE o del *Forum para los*

implementadores de esquemas de metadatos SCHEMAS. Es previsible que el FP6 (2002-2006) siga impulsando la investigación en metadatos en el contexto de las bibliotecas digitales y del acceso electrónico al patrimonio cultural europeo.

27. En relación con el proyecto de aplicación de metadatos elegido: una biblioteca digital real, editora de su colección, la *Biblioteca Virtual Miguel de Cervantes*, con el que finaliza esta tesis, podemos resumir las siguientes conclusiones:

- La BVC es un contexto finito (amplio, pero finito) de información, que cuenta en la actualidad con más de 6000 documentos electrónicos, y que contará al final del proyecto con 30.000 obras, digitalizadas y accesibles en la Web, de la literatura en lengua española.
- Actualmente, la recuperación de información en la BVC no se basa en metadatos sino en una base de datos bibliográfica constituida *a priori* sobre el documento original en papel.
- Los metadatos con los que cuenta actualmente son una sencilla y poco completa cabecera TEI extraída automáticamente a partir del catálogo bibliográfico realizado *a priori* y basado en el documento impreso.
- Creemos que lo que tiene que describir una biblioteca virtual son los DLOs que la constituyen y no el documento original impreso en el que se basan. Por ello, en nuestra propuesta introducimos un cambio en los flujos de trabajo de la producción y edición de los documentos digitales, de tal forma que los metadatos se creen vinculados al proceso de marcado en XML y edición digital. Así la cabecera de metainformación será la fuente de extracción de metadatos para la recuperación y no a la inversa.
- Proponemos para ello la utilización de un esquema RDF que soporte un perfil de aplicación del Dublin Core adaptado a la riqueza de contenidos de la BVC. De esta forma, se puede utilizar, por un lado, RDF como estructura de los metadatos para la recuperación de información en la propia biblioteca digital y, por otro, a través de XSLT, del mismo modo en que se convierten los

documentos marcados en TEI P4, según *cervante.dtd* en un documento HTML para la Web, se puede convertir la información codificada en RDF en metaetiquetas HTML con la semántica DC. Así se incrementará la calidad de la representación y recuperación de información en la BVC, a la vez que se utilizan las etiquetas <META> del HTML para la recuperación genérica en la Web. RDF, junto a otras tecnologías vinculadas a XML puede soportar todas las necesidades de metainformación de la Biblioteca Virtual *Miguel de Cervantes*.

- El componente de investigación, que subyace a la producción y digitalización y al proyecto de crear la mayor biblioteca virtual en lengua hispana, hace factible llevar a cabo el replanteamiento que proponemos en este trabajo para la creación de un sistema de recuperación de información basado en metadatos. Además, el proyecto de gestión de metadatos entra dentro de los objetivos del departamento de informática de la BVC para el año 2002, se han señalado además otras muchas fortalezas para ejecutarlo.

28. Por último, la investigación en metadatos y bibliotecas digitales está generando múltiples conceptos nuevos que tienen su origen en términos del inglés. Al igual que muchos informáticos han asumido la responsabilidad de discutir sobre la adaptación de la terminología informática en español, los profesionales de la información debemos asumir responsabilidades similares, relacionadas con la creación lingüística y la designación de nuevas realidades informativas, en el seno de los campos de trabajo vinculados a las tecnologías. (*Vid. Glosario*).

Con todo, esta investigación no termina aquí; constituye una pequeña aportación y una propuesta de aplicación al extremadamente complejo mundo de los metadatos, que nos llevarán hacia sistemas recuperación de información más coherentes. Como dijimos al principio, el pródigo mundo de la organización de la información en Internet, y particularmente de los metadatos, aún se encuentra en fase de desarrollo y construcción técnica, lo que nos hace prever que los profesionales de la información sólo estaremos limitados por la imaginación en la creación y aplicación de metainformación. Los bibliotecarios necesitamos encontrar un nuevo

papel en el inextricable e inabarcable mundo de la ubicuidad de la información electrónica, tratando de dar valor añadido a esa información sin tratar de aparentar un nivel de control que no podemos, ni es necesario, conseguir. Este trabajo se ampliará pues, en un futuro, a tenor del seguimiento de los avances y desarrollos que, sin duda, se seguirán produciendo en el ámbito de la metainformación. Si bien hasta aquí hemos reconocido la utilidad de los metadatos en entornos finitos de información, esperamos que, a tenor de un mayor esfuerzo normalizador y de la investigación en interoperabilidad, los metadatos —esa información estructurada para la información distribuida en Internet— sean también la clave de una Web Semántica y una de las piedras angulares del acceso universal a la información.

Queremos cerrar esta tesis retomando el apotegma de Eliot que citábamos en las primeras páginas: una pregunta retórica *¿dónde está el conocimiento que hemos perdido en información?* (*Where is the knowledge we have lost in information?*), que parece no tener respuesta... La posesión de información es una condición necesaria para la toma de decisiones y para la producción de conocimiento. Internet supone una hiper-inflación de información, un dédalo virtual de objetos de información donde sólo una recuperación de calidad: precisa, relevante y exhaustiva, puede proporcionar al hombre la información que necesita para competir en la nueva Sociedad de la Información, evitando quizás de esta forma, que su conocimiento se anegue en información. Por todo ello, se crean bibliotecas digitales (virtuales o electrónicas), por todo ello, surgen nuevas soluciones como los metadatos y por todo ello también, hemos realizado esta tesis doctoral y seguiremos investigando en este tema. Asimilando el pensamiento del lexicógrafo y poeta inglés del siglo XVIII, que también citamos al inicio de este trabajo, una *pregunta sólo da lugar a otra, un libro se refiere a otro libro*, de igual forma, una investigación como ésta dará lugar a otras.

BIBLIOGRAFÍA



NOTA:

Dada la inconsistencia e inestabilidad de los recursos Web, principal fuente de información en este trabajo, y a fin de dotar de mayor veracidad y actualidad a esta bibliografía en el momento de su presentación, todos los recursos electrónicos que se han manejado en el transcurso de la investigación se han revisado el **8-9 de diciembre de 2001**; por ello, no aparece fecha de consulta en este listado bibliográfico. No obstante, en caso de que los documentos referenciados hayan cambiado su ubicación y/o ya no estén operativos en el URL que se consigna, se ha mantenido la fecha original de consulta. Asimismo, ha cambiado la fecha de revisión y actualización (rev.) de algunas de estas fuentes, con respecto a las que aparecen citadas en el cuerpo de la tesis.

A

ABADAL FALGUERAS, Ernest. *Sistemas y servicios de información digital*. Gijón: Trea, 2001.

ABADAL FALGUERAS, Ernest, Assumció Estivil. *L'accés als recursos Web des de les biblioteques* [documento PDF]. Barcelona: Consorci de Biblioteques Univeritaries de Catalunya, gener de 2000. Disponible en: <http://www.cbuc.es/5digital/ireweb.pdf>

ABAITUA, Joseba. *Material de referencia para un curso de introducción a SGML* [documento HTML]. Bilbao: Universidad de Deusto, rev. 8 de mayo de 1997. Disponible en: <http://sirio.deusto.es/abaitua/konzeptu/sgml/sgml0.htm>

AGOSTI, M., F. Crivellari, M. Mclucci. The Effectiveness of Metadata and Other Content Descriptive Data in Web Information Retrieval [documento PDF]. En: *IEEE Meta-Data Conference (3. 1999. Bethesda)*. IEEE Computer Society, rev. 12 de febrero de 1999. Disponible en: <http://computer.org/conferen/proceed/meta/1999/papers/42/magosti.pdf>

AGULLO, Isidro. Internet invisible o Infranet: definición, clasificación y evaluación. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao)*. *La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p. 249-269.

AHMED, Kal, et al. *Professional XML Meta Data*. Birmingham: Wrox Press, 2001.

AHRONHEIM, Judy. Perspectives on... Descriptive Metadata: Emerging Standards. *Journal of Academic Librarianship*, September 1998, vol. 24, nº 5, p. 395-403.

Bibliografía

AHRONHEIM, Judy. *Judy and Magda's List of Metadata Initiatives*. [documento HTML]. University of Michigan, 2 de noviembre de 1997. Disponible en: <http://www-personal.umich.edu/~jaheim/alcts/bibaces.htm>

AHRONHEIM, Judy, et al. *TEI/MARC "Best Practices"* [documento HTML]. Ann Arbor: Michigan University Library, 16 de junio de 2001. Disponible en: <http://www.lib.umich.edu/staff/ocu/teiguide.html>

AKERROYD, John. La administración del cambio en las bibliotecas electrónicas. *IFLA Council and General Conference (66. 2000. Jerusalem)* [documento HTML]. IFLA, 1 de agosto de 2000. Disponible en: <http://www.ifla.org/IV/ifla66/papers/037-110s.htm>

ALBALA: *Sistema integrado de gestión de centros archivísticos* [documento PDF]. Madrid: Baratz, Servicios de Teledocumentación, 2000. Disponible en: <http://www.baratz.es/baratz/pdf/albala.pdf>

ALBERT, Ip, et al. Metasearching or Megasearchig: Toward a Data Model for Distributed Resource Discovery. En: *e-Education: Challenges and Opportunities. Hong Kong Web Symposium (5. 1999. Hong Kong)*, p. 65-82. También disponible en Internet en: <http://www.dls.au.com/metadata/DataModel.html>

ALDANA MONTES, J. F., M. M. Roldán García, A. C. Gómez Lora. Información semántica para la integración de servicios de datos sobre la Web. En: *ISKO (5. 2001. Alcalá de Henares). La representación y organización del conocimiento: Metodologías, modelos y aplicaciones* [CD-ROM]. Alcalá de Henares: Universidad de Alcalá de Henares, 2001, p 394-403..

ALEXANDER, Jan, Marsha Ann Tate. *Evaluating Web Resources* [documento HTML]. Wolfgram Memorial Library, Widener University, 8 de agosto de 1996, rev. 25 de julio de 2001. Disponible en: <http://muse.widener.edu/Wolfgram-Memorial-Library/webevaluation/webeval.htm>

ALONSO PIÑEIRO, Armando. Pobreza y riqueza del español en el tercer milenio [documento RTF]. En: *Congreso Internacional de la Lengua Española (2. 2001. Valladolid)*. Disponible en: http://cvc.cervantes.es/obref/congresos/valladolid/informacion/ponencia/pineiro_a.doc

ALSCHULER, Liora. *ABCD...SGML: A User's Guide to Structured Information*. London, etc.: International Thomson Computer Press, 1995.

ALURI, Rao, Aldasdair Kemp, John J.Boll. *Subject Analysis in Online Catalogs*. Englewood: Libraries Unlimited, 1991.

AMERICAN LIBRARY ASSOCIATION. *Committee on Cataloging: Description and Access. Task Force on Metadata: Summary Report June 1999* [documento HTML]. ALA, junio 1999, rev. 7 de agosto de 1999. Disponible en: <http://www.ala.org/alcts/organizations/ccs/ccda/tf-meta3.html> (consultado el 2 de septiembre de 1999).

Bibliografía

AMERICAN LIBRARY ASSOCIATION. *Committee on Cataloging: Description and Access. Task Force on Metadata and Cataloging Rules: Final Report* [documento HTML]. ALA, 3 de junio de 1998, rev. 4 de junio de 1998. Disponible en: <http://www.ala.org/alcts/organization/ccs/ccda/ta-tei2.htm> (consultado el 18 de junio de 1998).

ANDERSON, Martha. A Tool for Building Digital Libraries [documento HTML]. *D-Lib Magazine*, February 1999, vol. 5, nº 2. Disponible en: <http://www.dlib.org/dlib/february99/02journalreview.html>

ANDRESEN, Leif. Metadata in Denmark. *Vine (Theme issue: Metadata. Part 2)*, 1999, nº 117, p. 18-23.

ANGLADA i de Ferrer, Lluís M^a. Biblioteca Digital ¿mejor, peor o sólo distinto? *Anales de Documentación*, nº 3, 2000, p. 25-39. También disponible en Internet en: <http://www.um.es/fccd/anales/ad03/02bibliodigital.PDF> [Este documento apareció publicado en febrero de 1999 en: <http://www.cbuc.es/articulos/bdigupf/anglada.htm>].

ANGÓS ULLATE, José María, José Antonio Salvador Oliván, María Jesús Fernández Ruíz. Metadatos en documentos HTML: una ayuda para la recuperación de información. *Scire*, Julio-diciembre 1999, vol. 5, nº 2, p. 47-61.

ANSI/NISO Z39.50-2001. Information Retrieval (Z39.50): Application Service Definition and Protocol Specification [documento PDF]. Bethesda, Maryland: NISO, rev. 21 de noviembre de 2001. Disponible en: <http://www.loc.gov/z3950/agency/revision/part1.pdf> y <http://www.loc.gov/z3950/agency/revision/part2.pdf>

ANSI/NISO Z39.85-200x: The Dublin Core Metadata Element Set. Status: Reconsideration ballot February 2-March 5, 2001 [documento PDF]. Bethesda, Maryland: NISO, rev. 1 de febrero de 2001. Disponible en: <http://www.niso.org/pdfs/Z3985-REV.pdf>

ANSI/NISO Z39.85-2001: The Dublin Core Metadata Element Set: An American National Standard Developed by the National Information Standards Organization Approved September 10, 2001 by the American National Standards Institute. [documento PDF]. Bethesda, Maryland: NISO, rev. 12 de octubre de 2001. Disponible en: <http://www.niso.org/standards/resources/Z39-85.pdf>

APPLICATION of Terminology and Classification Tools for Digital Collection Development and Network-based Search: ACM DL Workshop (1998. Pittsburgh). [documento HTML]. Alexandria Digital Library. Santa Bárbara: University of California, rev. 13 de julio de 1998. Disponible en: http://www.alexandria.ucsb.edu/~lhill/nkos/dl98_workshop_preliminary.html (consultado el 28 de agosto de 1998).

APPLYING the AGLS Metadata Standard [documento HTML]. Canberra: Commonwealth of Australia, rev. 4 de Julio de 2000. Disponible en: http://www.naa.gov.au/recordkeeping/gov_online/agls/guidelines/contents.html

Bibliografía

APPS, Ann and Ross MacIntyre. *zetoc*: a Dublin Core Based Current Awareness Service [documento PDF]. *Journal of Digital Information*, 2001, vol. 2, issue 2. Disponible en: <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Apps/apps-v2.pdf>

ARMS, Caroline R. Some Observation on Metadata and Digital Libraries [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 29 de septiembre de 2000, 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/arms_paper.html

ASLAN, Goksel, Dennis McLeod. Semantic Heterogeneity Resolution in Federated Databases by Metadata Implantation and Stepwise Evolution. *The VLDB Journal*, 1999, n° 8, p. 120-132.

AURORA de Netscape [documento HTML]. Madrid, Barcelona: ADD, rev. 9 de julio de 1998. Disponible en: http://www.add.es/html/body_aurora_de_netscape.html (consultado el 24 de agosto de 1998). Este artículo está accesible ahora en: <http://home.netscape.com/es/newsref/pr/newsrelease501.html>

B

BACA, Murtha. A Picture Is Worth a Thousand Words: Metadata for Art Objects and Their Visual Surrogates. 2000, 15 p., (ejemplar cedido por la autora).

BAILEY, Charles W., Jr. *Scholarly Electronic Publishing Bibliography* [documento HTML]. Houston: University of Houston Libraries, 1996-98. Disponible en: <http://info.lib.uh.edu/sepbb/sepbb.html> (consultado el 18 de junio de 1998).

BAKER, Thomas. A Grammar of Dublin Core [documento HTML]. *D-Lib Magazine*, October 2000, vol. 6, n° 10. Disponible en: <http://www.dlib.org/dlib/october00/baker/10baker.html>

BAKER, Thomas, et al. What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives [documento PDF]. *Journal of Digital Information*, 2001, vol. 2, issue 2. Disponible en: <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Baker/baker-final.pdf>

BANERJEE, Kyle. Describing Remote Electronic Documents in the Online Catalog: Current Issue. *Cataloging and Classification Quarterly*, 1997, vol. 25, n° 1, p. 5-20.

BARSTOW, Art. Survey: RDF Data Stores [correo electrónico en lista de distribución]. *www-rdf-interest*, 19 de abril de 2001; 15:04. Disponible en: <http://lists.w3.org/Archives/Public/www-rdf-interest/2001Apr/0327.html>

BARU, Chaitanya, Arcot Rajasekar. A Hierarchical Access Control Scheme for Digital Libraries. En: *ACM Digital Libraries (3. 1998. Pittsburg)*. Ian Witten, Rob Akscyn, Frank M. Shipman, III, eds. New York: ACM, 1998, p. 275-276

Bibliografía

BEACOM, Matthew. Crossing a Digital Divide: AACR2 and Unaddressed Problems of Networked Resources [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 16 de octubre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/beacom_paper.html

BEAGLE, Donald. Visualization of Metadata. *Information Technology and Library*, December 1999, vol. 18, nº 4, p. 192-199. También disponible en: http://www.lita.org/ital/1804_beagle.html

BECK, Melissa. *CONSER Cataloging Manual: Module 31 Remote Access Computer File Serials* [documento HTML]. Washington, D.C: Library of Congress. Cooperative Online Serials, rev. 3 de diciembre de 2001. Disponible en: <http://lcweb.loc.gov/acq/conser/module31.html>

BECKETT, David. IAFA Templates in Use as Internet Metadata [documento HTML]. En: *International World Wide Web Conference (4. 1995. Boston)*, rev. 15 de diciembre de 1995. Disponible en: <http://www.w3.org/Conferences/WWW4/Papers/52>

BECKETT, David, Eric Miller. *Expressing Simple Dublin Core in RDF/XML* [documento XHTML]. DCMI, 28 de noviembre de 2001. Disponible en: <http://www.dublincore.org/documents/2001/11/28/dcmes-xml>

BEDFORD, Sarah. A Tangled WEB: An Industry Reinventing Itself, Internet Consulting Hunkers Down for the Long Haul [documento HTML]. *Business Forward*, May 2000, issue 25. Disponible en: <http://www.bizforward.com/wdc/archives/2001-05/buzzguide>

BERNERS-LEE, Tim. Metadata Architecture [documento HTML]. *Design Issues*. W3C, 6 de enero de 1997, rev. 21 de septiembre de 2000. Disponible en: <http://www.w3.org/DesignIssues/Metadata.html>

BERNERS-LEE, Tim. *Tejiendo la red: el inventor del World Wide Web nos descubre su origen*. Madrid: Siglo veintiuno, 2000.

BERNERS-LEE, Tim, James Hendler and Ora Lassila. The Semantic Web [documento HTML]. *Scientific American*, May 2001, issue 501. Disponible en: <http://www.scientificamerican.com/2001/0501issue/0501berners-lee.html>

BEYOND Book Indexing. Diane Brenner and Marilyn Rowland, eds. Medford: Information Today, The American Society of Indexers, 2000.

BIA, Alejandro. Technical Aspects of the Production Process of Digital Books Using XML-TEI at the Miguel de Cervantes Digital Library [documento HTML]. En: *International Conference of the Association for Computers and the Humanities and the Association for Literary and Linguistic Computing (2001. New York)*. New York:

Bibliografía

University, Humanities Computing Group, rev. 15 de mayo de 2001. Disponible en: http://www.nyu.edu/its/humanities/ach_allc2001/posters/bia/index.html

BIA, Alejandro. The Use of Multimediality to Enhance the Accessibility to Digital Library Resources: The Multicultural-scope of the Services Offered by the Miguel de Cervantes Digital Library Project. En: *Digital resources for the humanities: Conference abstracts*. London: University of London, SOAS, 2001, p. 9-14.

BIA, Alejandro, Andrés Pedreño. The Miguel de Cervantes Digital Library: The Hispanic Voice on the Web. *Literary and Linguistic Computing*, 2001, vol. 16, n° 1, p. 149-166.

BIA, Alejandro. Manuel Sánchez. A Reusable Hypermedia Design for DL Manuscripts, Based on XML, XSLT and Java. 2001, (ejemplar cedido por el autor).

BIA, Alejandro, Rafael C. Carrasco. Automatic DTD Simplification by Examples. En: *ACH/ALLC 2001: The Joint International Conference (2001. New York). Abstracts*, p. 7-9.

BIA, Alejandro, Rafael Muñoz. Aplicación de técnicas de extracción de información a bibliotecas digitales. En: *Congreso de la SEPLN (16. 2000. Vigo)*. Vigo: Universidad de Vigo, 2000, p. 207-214.

La Biblioteca Virtual Miguel de Cervantes [documento HTML]. Alicante: Biblioteca Virtual Miguel de Cervantes, Disponible en: <http://www.cervantesvirtual.com/proyectoES/BIMICESA03.shtml>

BIBLIOTECA Virtual Miguel de Cervantes: Dossier de Prensa [CD-ROM]. Universidad de Alicante; Banco Santander Central Hispano, [2001].

BIBLIOTECA Virtual Miguel de Cervantes: 2º aniversario. Universidad de Alicante; Banco Santander Central Hispano, [folleto distribuido el 27 de julio de 2001 con motivo del 2º aniversario de la BVC].

BIBLIOTHECA Universalis: a G-7 Global Information Society Pilot Project [documento HTML]. London: British Library, Gabriel, 1998. Disponible en: <http://www.bl.uk/gabriel/bibliotheca-universalis/index.html>

BIBLIOTHECA Universalis Project: Partners, Objectives and Achievements 1996-1998 [documento HTML]. London, etc.: Gateway to Europe's National Libraries, rev. 20 de junio de 2001. Disponible en: <http://portico.bl.uk/gabriel/bibliotheca-universalis/bibuniv.htm>

BIDE, Mark & Associates. *Standards for Electronic Publishing: An Overview* [documento PDF]. NEDLIB Project. La Hague: National Library of Netherlands, Koninklijke Bibliotheek, agosto de 2000. Disponible en: <http://www.kb.nl/coop/nedlib/results/e-publishingstandards.pdf>

Bibliografía

BISHOFF, Liz, William A. Garrison. Metadata, Cataloging, Digitization and Retrieval: Who's Doing What to Whom: The Colorado Digitization Project Experience [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 7 de noviembre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/bishoff_paper.html

BLAIR, David C. *Language and Representation in Information Retrieval*. Amsterdam, etc.: Elsevier Science Publishers, 1990, 335 p.

BLUE ANGEL TECHNOLOGIES. Metadata Explained [documento HTML]. *News and Info*, 1996-1998, rev. 10 de abril de 1998. Disponible en: <http://www.bluelangel.com/NewsAndInfo/faqs/metadata.htm> (consultado el 19 de junio de 1998).

BOWMAN, Mic, et al. *The Harvest Information Discovery and Access System*. Harvest, 2 de septiembre de 1996. Disponible en: <http://harvest.transarc.com> (consultado el 8 de julio de 1998).

BONILLA ÁLVAREZ, Sebastián. Información y relevancia. Hipótesis acerca de cómo procesamos los seres humanos la información lingüística. En: *Revista Española de Documentación Científica*, 1996, vol. 19, nº 4, p. 392-410.

BORDEN, Jonathan. *Terminology Definition Description Language (TDDL) 1.0* [documento XHTML]. Open Healthcare Group, rev. 23 de marzo de 2001. Disponible en: <http://www.openhealth.org/RDDL/tddl>

BORGMAN, Christine L. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World*. Cambridge, London: The MIT Press, 2000.

BORGMAN, Christine L. Where is the Librarian in the Digital library? *Communications on the ACM*, May 2001, vol. 44, nº 4, p. 66-67.

BOSAK, Jon, Tim Bray. XML and the Second-Generation Web [documento HTML]. *Scientific American*, May 1999. Disponible en: <http://www.sciam.com/1999/0599issue/0599bosak.html>

BOURDEAU, Robert H., et al. Institutions, Languages, Content Standards, Record Syntaxes, and Protocols: Fitting it all Together [documento HTML]. En: *IEEE Metadata Conference (2. 1997. Silver Spring)*. IEEE Computer Society, 1997, rev. 17 de febrero de 1998. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/rbourdeau/rbourdeau.html>

BRACK, Verity. The eLib Collection Description Schema. *Vine (Theme issue: Metadata. Part 1)*, 1999, nº 116, p. 38-48.

Bibliografía

BRADLEY, Phil. Meta tags – What, Where, When, Why? [documento HTML]. Phil Bradley Home Page, 14 de mayo de 1999. Disponible en: <http://www.philb.com/metatag.htm>

BRADLEY, Phil. The Relevance of Underpants to Searching the Web [documento HTML]. *Ariadne*, 23-Jun-2000, issue 24. Disponible en: <http://www.ariadne.ac.uk/issue24/search-engines>

BRADLEY, Phil. Virtual Libraries and Internet Searching. *Online & CD-ROM Review*, 1999, vol. 23, nº 6, p. 353-355.

BRAY, Tim. Beyond HTML: XML and Automated Web Processing [documento HTML]. *View Source Magazine*. Netscape Communications Corporation, 1999. Disponible en: http://developer.netscape.com/viewsource/bray_xml.html, y en: <http://xml.coverpages.org/brayXMLViewsource.html>

BRAY, Tim. RDF and Metadata [documento HTML] *XML.com*, Junio de 1998. Disponible en: <http://www.xml.com/xml/pub/98/06/rdf.htm> (consultado el 17 de julio de 1998).

BRAY, Tim. What is RDF? [documento HTML] [Actualización de RDF and Metadata <<http://www.xml.com/xml/pub/98/06/rdf.htm>> realizada por Dan Brickley]. *XML.com*, 24 de enero de 2001. Disponible en: <http://www.xml.com/pub/a/2001/01/24/rdf.html>

BRASETHVIK, Terje. A Semantic Modeling Approach to Metadata. *Internet Research: Electronic Networking Applications and Policy*, 1998, vol. 8, nº 5, p. 377-386.

BRICKLEY, Daniel. Mozilla RDF / Z39.50 Integration Project [documento HTML]. The Mozilla Organization, 22 de agosto de 1999. Disponible en: <http://www.mozilla.org/rdf/doc/z3950.html>

BRICKLEY, Daniel (daniel.brickey@bristol.ac.uk). ANNOUNCE: Mozilla RDF/Z39.50 Integration Project [mensaje en newsgroup]. *Netscape.public.mozilla.rdf*, 24 de agosto de 1999 01:04:50 GMT

BRICKLEY, Dan, Eric Miller. A Story about RDF and XML. XML Processing Workshop Position Paper, June 2001 [documento HTML]. W3C, 26 de junio de 2001. Disponible en: <http://www.w3.org/2001/06/rdf-xproc/1>

BRISSON, Roger. The World Discovers Cataloging: A Conceptual Introduction to Digital Libraries, Metadata and the Implications for Library Administrations. *Journal of Internet Cataloging*, 1999, vol. 1, nº 4, p. 3-30.

BRONSOILER FRID, Alfredo. De sistemas de bibliotecas a sistemas de información [documento PDF]. En: *Interfaces 2000: Foro internacional sobre biblioteca digital (2000. Colima)*. Disponible en: <http://www.ucol.mx/interfaces/interfaces2000/mesast/Mt01.pdf>

Bibliografía

BUCKLAN, Michael. Information as a Thing. *Journal of the American Society for Information Science*, June 1991, vol. 42, n° 5, p. 351-360. (El preprint de este artículo, disponible en Internet en: <http://www.dlib.org/dlib/january99/buckland/01buckland.html>).

BUCKLAN, Michael, et al. Mapping Entry Vocabulary to Unfamiliar Metadata Vocabularies [documento HTML]. *D-Lib Magazine*, January 1999, vol. 5, n° 1. Disponible en: <http://www.dlib.org/dlib/january99/buckland/01buckland.html>

BURNARD, Lou. *La TEI simplifiée: une introduction au codage des textes électroniques en vue de leur échange* [documento HTML]. Chicago: University of Illinois, 24 de junio de 1996. Disponible en: http://www.uic.edu/orgs/tei/lite/teiu5_fr.html (consultado el 6 de junio de 1998).

BURNARD, Lou. TEI and XML: a Marriage Made in Heaven. An Introduction to the Future of Digital Information [documento HTML]. En: *Computing Arts: Digital Resources for Research in the Humanities (2001. Sydney)*. University of Sydney, Research Institute for Humanities & Social Sciences, octubre 2001. Disponible en: <http://www.arts.usyd.edu.au/Arts/departs/rihss/keyabs.html#Burnard>

BURNARD, Lou. What is SGML and How Does It Help? *Computers and the Humanities*, 1995, vol. 29, p. 41-50.

BURNARD, Lou. Text Encoding for Interchange: a New Consortium [documento HTML]. *Ariadne*, 23-Jun-2000, issue 24. Disponible en: <http://www.ariadne.ac.uk/issue24/tei>

BURNARD, Lou, Richard Light. Three SGML Metadata Formats: TEI, EAD, and CIMI [documento HTML]. En: *Work Package 1 of Telematics for Libraries project BIBLINK (LB4034)*. UKOLN Metadata Group, diciembre de 1996, rev. 14 de mayo de 1998. Disponible en: <http://www.ukoln.ac.uk/metadata/BIBLINK/wp1/sgml>

BURNARD, Lou, C. M. Sperberg-McQueen. *TEI Lite: Una introducción al Text Encoding for Interchange*, Junio de 1995. [documento HTML]. Traducción de Manuel Sánchez Quero. Alicante: Universidad de Alicante, rev. 5 de noviembre de 2001. Disponible en: http://www.alu.ua.es/m/msq1/Trad_TEI-Lite.htm

BURNETT, Kathleen, Kwong Bor Ng, Soyeon Park. A Comparison of the Two Traditions of Metadata Development. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, n° 13, p. 1209-1217.

BUTTERFIELD, Kevin L. Cataloger's and the Creation of Metadata Systems: A Collaborative Vision at the University of Michigan [documento HTML]. En: *OCLC Internet Cataloging Colloquium (1996. San Antonio)*. OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/butter.htm>

C

CAMPBELL, Debbie. The Creation and Use of the Australian MetaMatters Web Site. *Vine (Theme issue: Metadata. Part 2)*, 1999, n° 117, p. 49-53.

CAPLAN, Priscila. International Metadata Initiatives: Lessons in Bibliographic Control [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 1 de noviembre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/caplan_paper.html

CAPLAN, Priscilla. You Call it Corn, We Call it Syntax-independent Metadata for Document-Like Objects [documento HTML]. *The Public-Access Computer Systems Review*, 1995, vol. 6, n° 4. Disponible en: <http://info.lib.uh.edu/pr/v6/n4/capl6n4.html> y en: <http://www.nlc-bn.ca/ifla/documents/libraries/cataloging/caplan3.txt> (consultado el 15 de junio de 1998).

CAPLAN, Priscilla, et al. *Report on the Break-out Group on Standardisation DC-6, Washington, DC*; 3 November 1998 [documento XHTML]. DCMI, 28 de febrero de 2001. Disponible en: <http://dublincore.org/workshops/dc6/pp/standardisation.shtml>

CAPURRO, Rafael. *Las bibliotecas en la Era Digital* [documento HTML]. Capurro.de, rev. 6 de diciembre de 2000. Disponible en: <http://www.capurro.de/gohete.htm>

CAPURRO, Rafael. *Epistemology and Information Science. REPORT TRITA-LIB-6023*. Stephan Schwarz, ed. August 1985 [documento HTML]. Disponible en: <http://www.capurro.de/trita.htm>

CASTRO, Elizabeth. *Guía de aprendizaje XML*. Madrid, etc.: Prentice Hall, 2001.

CATALOGING Internet Resources: A Manual and Practical Guide [documento HTML], Nancy B. Olson, ed. 2nd ed. OCLC, 1997, rev. 28 de julio de 1998. Disponible en: <http://www.oclc.org/oclc/man/9256cat/toc.htm>

CATALOGING Guidelines for Internet Resources. Drafted by Electronic Publications Implementations Team [documento HTML]. New York: University of Rochester Libraries, septiembre 1997, rev. 18 de septiembre de 1998. Disponible en: <http://www.lib.rochester.edu/cat/catguide.htm>

CATHRO, Warwick. *Metadata: An Overview* [documento HTML]. Canberra: National Library of Australia, rev. 10 de septiembre de 1997. Disponible en: <http://www.nla.gov.au/nla/staffpaper/cathro3.html>

CATLEDGE, L.D. and J.E. Pitkow. Characterizing Browsing Strategies in the World-Wide Web [documento HTML]. En: *International World-Wide Web Conference (3. 1995. Darmstadt)*. Computer Graphics in Darmstadt: Fraunhofer-Institut für

Bibliografía

Graphische Datenverarbeitung, abril de 1995. Disponible en: <http://www.igd.fhg.de/www/www95/proceedings/papers/80/userpatterns/UserPatterns.Paper4.formatted.html>

CAUSTON, Laurie. *Identifying and Describing Web Resources* [documento PDF]. El Pub: Electronic Publishing, 17 de noviembre de 1998. Disponible en: http://www.elpub.org/assets/images/web_resources.pdf (consultado el 10 de octubre de 2000).

CHANG, May. An Electronic Finding Aid Using Extensible Markup Language (XML) and Encoded Archival Description (EAD). *Library Hi Tech*, 2000, vol. 18, n° 1, p. 15-27.

CHANG, Wendy and Aindong Zhang. Metadata for Distributed Visual Database Access [documento HTML]. En: *IEEE Meta-data Conference (2. 1997. Silver Spring)*. IEEE Computer Society, 1997, rev. 17 de febrero de 1998. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/azhang/azhang.html>

CHAPMAN, Ann, Michael Day and Debra Hiom. Metadata: Cataloguing Practice and Internet Subject-based Information Gateways [documento HTML]. *Ariadne*, 10-December-1998, issue 18. Disponible en: <http://www.ariadne.ac.uk/issue18/metadata>

CHEN, Chao-chen, Hsueh-hua Chen, Kuang-hua Chen. The Design of Metadata Interchange for Chinese Information and Implementation of Metadata Management System. *Bulletin of American Society for Information Science and Technology*, June/July 2001, vol. 27, n° 5, p. 21-27.

CHEN, Hsinchun. Semantic Research for Digital Libraries [documento HTML]. *D-Lib Magazine*, October 1999, vol. 5, n° 10,. Disponible en: <http://www.dlib.org/dlib/october99/chen/10chen.html>

CHEN, Hsinchun, et al. Internet Browsing and Searching: user Evaluations of Category Map and Concept Space Techniques. *Journal of the American Society of Information Science*, May 1998, vol. 49, n° 7, p. 582-603.

CHEN, Ya-ning, et al. A Case Study in Designing Chinese Metadata. *Online Information Review*, 2000, vol. 24, n° 3, p. 229-233.

CHEONG, Fah-Chun. *Internet Agents: Spiders, Wanderers, Brokers and Bots*. Indianápolis : New Rider Publishing, 1996.

CHESNUTT, David R. The Model Editions Partnership: "Smart Text" and Beyond [documento HTML]. *D-Lib Magazine*, July/August 1997, vol. 3, n° 7/8. Disponible en: <http://www.dlib.org/dlib/july97/07chesnutt.html>

CHOWDHURY, G. G., Sudatta Chowdhury. Digital Library Research: Major Issues and Trends. *Journal of Documentation*, September 1999, vol. 55, n° 4, p. 409-448.

Bibliografía

CHOWDHURY, G. G. *Introduction to Modern Information Retrieval*. London: Library Association Publishing, 1999.

CLARK, James. *Comparison of SGML and XML*. *World Wide Web Consortium Note 15-December-1997* [documento HTML]. W3C, 15 de diciembre de 1997. Disponible en: <http://www.w3.org/TR/NOTE-sgml-xml.html>

CLARK, James. *Beyond the Dublin Core: Rich Meta-Data and Convenience-of Use Are Compatible After All* [documento HTML]. Canberra: Australian National University, 11 July, 1997. Disponible en: <http://www.anu.edu.au/people/Roger.Clarke/II/DublinCore.html>

CLEARINGHOUSE de datos geográficos de El Salvador: preguntas frecuentes [documento HTML]. San Salvador: Ministerio de Medio Ambiente, rev. 23 de abril de 2001. Disponible en: <http://www.marn.gob.sv/clearinghouse/menufoot/preg.htm> (consultado el 2 de junio de 2001).

CODINA, Lluís. Cómo funcionan los servicios de búsqueda en Internet: un informe especial para navegantes y creadores de información (Parte I). *IWE*, mayo 1997, vol. 6, nº 5, p. 22-27.

CODINA, Lluís. Evaluación de recursos digitales en línea: conceptos, indicadores y métodos. *Revista Española de Documentación Científica*, 2000, vol. 23, nº 1, p. 9-44.

COLIN, Smith. *Collins: Diccionario español-inglés inglés-español*. 5ª ed. Barcelona: Grijalbo Mondadori, 1997.

COMPUTERS in libraries: North America's Largest Technology Conference & Exhibition for Librarians and Information Managers (16. 2001. Washington). Carol Nixon, comp. Medford: Information Today, 2001.

CONTRIBUTING Information to CERES: Metadata [documento HTML]. California Environmental Resources Evaluation System, California Resources Agency, 1996, rev. 31 de marzo de 1997. Disponible en: <http://ceres.ca.gov/help.desk/metadata.html>

COOMBS, James H., Allen H. Renear and Steven J. DeRose. Markup Systems and the Future of Scholarly Text Processing. *Communications of the ACM*, 1987, vol. 30, issue 11, p. 933-947. También disponible en Internet en: <http://www.oasis-open.org/cover/coombs.html>

The CORE Categories for Visual Resources: Version 2.0, October 15, 1997 [documento HTML]. Visual Resources Association, Data Standards Committee, rev. 20 de octubre de 1997. Disponible en: <http://www.oberlin.edu/~art/vra/guide.html> (consultado el 24 de agosto de 1998).

Bibliografía

CORNELLÁ, Alfons. La importancia de la relevancia en información. [correo electrónico en lista de distribución]. *Extra-Net!*, 20 de enero de 1998; 00:18. Disponible en: <http://www.infonomia.com/extranet/index.asp?idm=1&idrev=1&num=305>

CORTEZ, Edwin M. Use of Metadata Vocabularies in Data Retrieval. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, n° 13, p. 1218-1223.

COVER, Robin. XML and Semantic Transparency [documento HTML]. En: *The XML Cover Pages*. Oasis, 24 de noviembre de 1998. Disponible en: <http://www.oasis-open.org/cover/xmlAndSemantics.html>

COVERT, Kay. CORC: Helping Libraries Take a Leading Role in the Digital Age. En: *European Conference on Research and Advanced Technology for Digital Libraries (4. 2000. Lisbon): ECDL2000*. New York, etc.: Springer, 2000, p. 399-402.

COX, Graeme. *Regarding your Klarity Enquiry* [correo electrónico personal], 28 de marzo de 2001; 02:33. Disponible en: emendez@bib.uc3m.es

CREATING Web-Accessible Databases: Case Studies for Libraries, Museums and Other Nonprofits. Julie M. Still, ed. Medford: Information Today, 2001.

CRAVEN, Timothy C. Changes in Metatag Descriptions Over Time [documento HTML]. *First Monday*, October 2001, vol. 6, n° 10. Disponible en: http://www.firstmonday.dk/issues/issue6_10/craven/index.html

CRAVEN, Timothy C. DESCRIPTION Meta Tags in Public Home and Linked Pages [documento HTML]. *LIBRES: Library and Information Science Research*, September 2001, vol. 11, issue 2. Disponible en: <http://libres.curtin.edu.au/LIBRE11N2/craven.htm>

CROMWELL-KESSLER, Willy and Ricky Erway. *Metadata Summit (1997. Mountain View): Meeting report* [documento HTML]. Research Library Group, rev. 7 de julio de 1997. Disponible en: <http://www.rlg.org/meta9707.html>

CROSS, Phil, Dan Brickley, Traugott Koch. *RDF Thesaurus Specification (draft)* [documento XHTML]. ILRT, 24 de enero de 2000, rev. 24 de enero de 2001. Disponible en: <http://ilrt.org/discovery/2001/01/rdf-thes>

CRUZ RUISECO, Yanet. Los metadatos: seguridad en la recuperación de los recursos de Internet. En: *Congreso Internacional de Información: INFO99 (1999. La Habana)* [CD-ROM]. La Habana: IDICT, 1999.

D

DAHAN, Michael. Counting Angels on a Pinhead: Critically Interpreting Web Size Estimates. *Online*, Jan./Feb. 2000, p. 35-40. También disponible en Internet en: <http://www.onlineinc.com/onlinemag/OL2000/dahn1.html>

DAY, Michael, Rachel Heery and Andy Powell. National Bibliographic Records in the Digital Information Environment: Metadata, Links and Standards. *Journal of Documentation*, January 1999, vol. 55, nº 1, p. 16-32.

DC-2001. *Proceedings of the International Conference on Dublin Core and Metadata Applications* [documento PDF]. Keizo Oyama and Hironobu Gotoda, eds. Tokyo: National Institute of Informatics, 2001. Disponible en: <http://www.nii.ac.jp/dc2001/proceedings/product/book.pdf>

DEACON, Prue. Simplicity vs. Structure: Which Way for the Dublin Core? *Cataloguing Australia*, 1999, vol. 25, nº 1/4, p. 36-43.

DECISION Nº 182/1999/EC of the European Parliament and of the Council of 22 December 1998 Concerning the Fifth Framework Programme of the European Community for Research, Technological Development and Demonstration Activities (1998 to 2002) [documento HTML]. Cordis, 1999. Disponible en: <http://www.cordis.lu/fp5/src/ec-en5.htm>

DEKKERS, Makx, Christian Eilert. *Standards Framework Report 2* [documento PDF]. SCHEMAS, 27 de Mayo 2001. Disponible en: <http://www.schemas-forum.org/stds-framework/2.pdf>

DELSEY, Tom. The Library Catalogue in a Networked Environment [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 16 de julio de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/delsey_paper.html

DEMPSEY, Lorcan. Meta Detectors [documento HTML]. *Ariadne*, 20-May-1996, issue 3. Disponible en: <http://www.ariadne.ac.uk/issue3/metadata>

DEMPSEY, Lorcan. Roads to Desire: Some UK and Other European Metadata and Resource Discovery Projects [documento HTML]. *D-Lib Magazine*, July/August 1996, vol. 2, nº 7/8. Disponible en: <http://www.dlib.org/dlib/july96/07dempsey.html>

DEMPSEY, Lorcan and Rachel Heery. Metadata: A Current View of Practice and Issues. *Journal of Documentation*, March 1998, vol. 54, nº 2, p. 145-172.

DEMPSEY, Lorcan and Rachel Heery. *A Review of Metadata: A Survey of Current Resource Description Formats* [documento RTF]. Bath: University, UKOLN

Bibliografía

Metadata Group, 21 de marzo de 1997 (versión 1.0), rev. 28 de enero de 1999, 106 p. Disponible en: <http://www.ukoln.ac.uk/metadata/desire/overview/overview.rtf>

DENENBERG, Ray. *Structuring and Indexing the Internet* [documento HTML]. Washington, DC.: Library of Congress, 30 de diciembre de 1996. Disponible en: <http://lcweb.loc.gov/z3950/agency/papers/italy.html>

DESRICHARD, Yves. Les formats et normes de catalogage: évolution et perspectives. *Bulletin des Bibliothèques de France*, 1998, vol. 43, n° 3, p. 56-65. También disponible en Internet en: <http://www.enssib.fr/bbf/bbf-98-3/11-desrichard.pdf>

DESAI, Bipin C. *The Semantic Header and Indexing and Searching on the Internet* [documento HTML]. Montreal: Concordia University, Department of Computer Science, rev. 11 de agosto de 1995. Disponible en: <http://www.cs.concordia.ca/~faculty/bcdesai/cindi-system-1.0.html>

DESAI, Bipin C. *The Semantic Header and Indexing and Searching on the Internet* [documento HTML]. Montreal: University of Concordia, Department of Computer Science, febrero de 1995, rev. 30 de septiembre de 1997. Disponible en: <http://www.cs.concordia.ca/~faculty/bcdesai/cindi-system-1.1.html>

DESAI, Bipin C. Supporting Discovery in Virtual Libraries. *Journal of the American Society for Information Science*, 1997, vol. 48, n° 3, p. 190-204.

DESIRE. *Information Gateways Handbook (Print Version)* [documento HTML]. DESIRE, rev. 26 de abril de 2000. Disponible en: <http://www.desire.org/handbook/print4.html>

DAHAN, Michael. Counting Angels on a Pinhead: Critically Interpreting Web Size Estimates. *Online*, Janvier 2000., p. 35-40. También disponible en Internet en: <http://www.onlineinc.com/onlinemag/OL2000/dahn1.html>

DICCIONARIO de la lengua española. 21ª ed. Madrid: Real Academia Española, 1992, Tomos I y II.

DICCIONARIO de la lengua española: edición electrónica. Versión 21.1.0. [CD-ROM]. Madrid: Espasa-Calpe, 1995.

DICCIONARIO de la lengua española [documento HTML]. 22ª ed. Madrid: Real Academia Española, 8 de octubre de 2001. Disponible en: <http://buscon.rae.es/drae/drae.htm>

A DICTIONARY of HTML Metatags [documento HTML]. Vancouver Webpages, rev. 6 de diciembre de 2001. Disponible en: <http://vancouver-webpages.com/META>

DIGITAL IMAGING GROUP. *The Power of Metadata Is Propelling Digital Imaging Beyond the Limitations of Conventional Photography* [documento PDF]. DIG,

Bibliografía

Agosto de 1999. Disponible en: <http://members.digitalimaging.org/sheredocs/DIG35WhitePaper.pdf> (consultado el 18 de mayo de 2000).

DIGITAL IMAGING GROUP. *DIG35 Specification: Metadata for Digital Images. Version 1.0* [documento PDF]. DIG, 30 de agosto de 2000. Disponible en: <http://members.digitalimaging.org/shreddocs/downloads/dig35v1.0-sept00.pdf> (consultado el 18 de mayo de 2001).

DIGITAL Libraries: Philosophies, Technical Design, Considerations and Example Scenarios. David Stern, ed. New York, London, Oxford: The Haworth Press, 1999.

DILLON, Martin. Metadata for Web Resources: How Metadata Works on the Web [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 31 de agosto de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/dillon_paper.html

DIRECTORY Interchange Format (DIF) Writer's Guide, Version 7, 1999 [documento HTML]. Global Change Master Directory. National Aeronautics and Space Administration, May 1999. Disponible en: <http://gcmd.gsfc.nasa.gov/md/difguide/difman.html> (consultado el 10 de febrero de 2001).

DISCOVERING Online Resources Across the Humanities: A Practical Implementation of the Dublin Core. Paul Miller and Daniel Greenstein, eds. Bath: UKOLN, 1997, 95 p. También disponible en Internet en: http://ahds.ac.uk/public/metadata/disc_01.html (consultado el 22 de agosto de 2000).

DISTRIBUTED Indexing/Searching Workshop (1996. Cambridge). World Wide Web Consortium, 20 de junio de 1996, rev. 9 de julio de 1996. Disponible en: <http://www.w3.org/Search/9605-Indexing-Workshop/Papers/AllPaps.doc>

DOCUMENTACIÓN. Referencias bibliográficas. Contenido, forma y estructura: ISO 690-1987, PNE 50-104. *Revista Española de Documentación Científica*, 1992, vol. 15, nº 3, p. 281-288; 1993, vol. 16, nº 1, p. 53-68.

DORAN, Kelly. Metadata for a Corporate Intranet. *Online*, January/February, 1999, p.43-50.

DORNFEST, Rael . Writing RSS 1.0 [documento HTML]. *O'Reilly Network*, 25 de agosto de 2000. Disponible en: http://www.oreillynet.com/pub/a/network/2000/08/25/magazine/rss_tut.html

DOVEY, Matthew J. "Stuff" about "Stuff"- the Differing Meanings of "Metadata". *Vine (Theme issue: Metadata. Part 1)*, 1999, nº 116, p. 6-13.

DOVEY, Matthew J. UKOLN RDF Seminar, Bath [documento HTML]. *Ariadne*, 17-May-1998, issue 15. Disponible en: <http://www.ariadne.ac.uk/issue15/events/stakis.html>

Bibliografía

DRAFT Thesaurus of Information Object Type Terminology [documento HTML]. Alexandria Digital Library. Santa Bárbara: University of California, 6 de agosto de 1998. Disponible en: http://www.alexandria.ucsb.edu/~lhill/MultiTes_files/objtype.tes/index.htm

DUBLIN Core Projects[documento XHTML]. Dublin Core Metadata Initiative, Disponible en: <http://dublincore.org/projects>

DUCLOY, Jacques, Catherine Lupovici, Elizabeth Cherhal. Les metadonnées et le catalogage des documents numériques [documento HTML]. En: *Minutes des rencontres BIBLIO-FR (1998. Caen)*. Caen: Université de Caen, Laboratoire GREYC, abril de 1998, rev. 15 de enero de 1999. Disponible en: <http://www.info.unicaen.fr/bnum/biblio-fr/rencontres98/minutes/metadonnees/texte.html>

DUMBILL, Edd. RDF - Why We Should Care - and RSS [documento HTML]. *Edd Dumbill Weblog*. O'Reilly, 6 de marzo de 2000. Disponible en: [http://weblogs.oreillynet.com/edd/discuss/msgReader\\$65](http://weblogs.oreillynet.com/edd/discuss/msgReader$65) (consultado el 8 de julio de 2001).

DUMBILL, Edd. The Semantic Web: A Primer [documento HTML]. *XML.com*, 1 de noviembre de 2000. Disponible en: <http://www.xml.com/pub/2000/11/01/semanticweb>

E

EDUCOM. *Industry Consensus Reached on Labeling of Education Materials on the Internet: Educom's IMS Project "Meta-data" Specification to be Used in Software Tools and Content* [documento HTML]. Denis Newman, dir. San José: Imsproject, 23 de marzo de 1998. Disponible en: <http://www.imsproject.org/metadate/metadatapr2.html> (consultado el 17 de julio de 1998).

EDUCOM. IMS Meta-data [documento HTML]. En: *IMSProject*. Instructional Management Systems, rev. 3 de diciembre de 2001. Disponible en: <http://www.imsproject.org/metadate>

EILERT, Christian, Makx Dekkers. *Metadata Watch Report #6 and Standards Framework Report #3* [documento PDF]. SCHEMAS, 11 de noviembre de 2001. Disponible en: <http://www.schemas-forum.org/metadata-watch/6.pdf>

EITO BRUN, Ricardo. Colecciones digitales: análisis de usabilidad. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao)*. *La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p. 435-444.

EL ZAÏM, Adel et Sylvie Tellier. L'avenir des formats de communication: HTML, SGML et formats bibliographiques: de l'information à la métainformation [documento HTML]. En: *International conference on The Future of Communication*

Bibliografía

Formats (1996. Ottawa).. Hull (Canadá): Banque International d'information sur les États francophones, rev. 23 de febrero de 1996. Disponible en: <http://www.acctbief.org/avenir/crim.htm>

ELLERMAN, Castedo. *Channel Definition Format. Submitted to de W3C on 09 March 1997* [documento HTML]. W3C, 3 de octubre de 1997. Disponible en: <http://www.w3.org/TR/NOTE-CDFsubmit.html>

ELLIS, David and Nigel Ford. In Search of the Unknown User: Indexing, Hypertext and the World Wide Web. *Journal of Documentation*, January 1998, vol. 54, n° 1, p. 28-47.

ELLIOTT LYNN. *The ABC's of Evaluating WWW Sources* [documento HTML]. Panama City: Gulf Coast Community College Library, rev. 21 de febrero de 2001. Disponible en: <http://www.gc.cc.fl.us/library/researchhelp/webeval.htm> (consultado el 20 de mayo de 2001).

EMACS (specifically NTEmacs 19.34 for Windows 95/NT) [documento HTML]. En: *Ebenezer's software suite for TEI*. Manitoba: University of Manitoba, Linguistics Department, 27 de junio de 1997, rev. 3 de julio de 1997. Disponible en: <http://www.umanitoba.ca/faculties/arts/linguistics/russell/ebenezer.htm#emacs>

ENCISO CARVAJAL, Bertha. *La biblioteca: bibliosistemática e información*. 2ª ed. México: Colegio de México, 1997 [documento HTML]. Alicante: Biblioteca Virtual Cervantes, 1999-2001. Disponible en: <http://www.cervantesvirtual.com/servlet/SirveObras/079596188953519752324333>

ENRIQUEZ HARRIS, Pita. In Search of the Rosetta Stone. *Online & CD-ROM Review*, 1999, vol. 23, n° 4, p. 235-238.

ENSOR, Pat. *Tool Kit for the Expert Web Searcher* [documento HTML]. LITA, Top Technology Trends Committee; ALA, rev. 24 de noviembre de 2001. Disponible en: <http://www.lita.org/committe/toptech/toolkit.htm>

ERCEGOVAC, Zorana. Introduction. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, n° 13, p. 1165-1168.

ERSHOVA, Tatiana V. & Yuri E. Hohlov. Migrar de la biblioteca de hoy a la biblioteca de mañana: ¿Re-o E-volución? En: *IFLA Council and General Conference (66. 2000. Jerusalem)* [documento HTML]. IFLA, 29 de julio de 2000. Disponible en: <http://www.ifla.org/IV/ifla66/papers/063-110s.htm>

ESPELT, Costança. Improving Subject Retrieval: User-friendly Interfaces and Effectiveness [documento HTML]. *bid: Biblioteconomia i Documentació*, juny 1998, n° 1. Disponible en: <http://www.ub.es/biblio/bid/01espel1.htm>

Bibliografía

ESTIVILL RIUS, Assumpció. Organizació dels recursos Internet. *Item*, nº 18, 1996, p. 42-74.

ESTIVILL RIUS, Assumpció. El processament dels recursos electrònics [documento HTML]. *bid: Biblioteconomia i Documentació*, nº 1, juny 1998. Disponible en: <http://www.ub.es/biblio/bid/01estiv1.htm>

ESTIVILL RIUS, Assumpció, et al. Recursos web i metadades: informe del projecte [documento HTML]. *bid: Biblioteconomia i Documentació*, nº 7, desembre 2001. Disponible en: <http://www.ub.es/biblio/bid/07estiv1.htm>

EXCERPTS from International Standard ISO 690-2 Information and Documentation: Bibliographic References. Part 2: Electronic Documents or Parts Thereof [documento HTML]. ISO, 2000, rev. 14 de septiembre de 2000. Disponible en: <http://www.nlc-bnc.ca/iso/tc46sc9/standard/690-2e.htm>

EYRE, John. Distributed Image Services (Working Together). *Vine (Theme issue: Digital Images in Libraries)*, July 1998, nº 107, p. 65-72.

F

FABA PÉREZ, Cristina, Félix de Moya Anegón. Bibliotecas digitales: concepto y principales proyectos. *Investigación Bibliotecológica*, enero/junio 1999, vol. 13, nº 26, p. 64-78.

FAUSEY, Jon and Keith Shafer. All My Data is in SGML. Now What? *Journal of the American Society for Information Science*, 1997, vol. 7, nº 48, p. 638-643.

FEDERAL Geographic Data Committee. *Content Standard for Digital Geospatial Metadata (CSDGM)* [documento HTML]. Reston: FGDC, USGS, rev. 26 de junio de 2001. Disponible en: <http://www.fgdc.gov/metadata/constan.html>

FERNÁNDEZ CALVO, Rafael. *Glosario básico Inglés-Español para usuarios de Internet*. 4ª ed. Barcelona: Asociación de Técnicos de Informática, 2001. También disponible en Internet en: <http://www.ati.es/PUBLICACIONES/novatica/glointv2.html>

FIAB. *Descripción bibliográfica internacional normalizada para archivos de ordenador: ISBD (CF)*. Madrid: ANABAD, Arco-Libros, 1993.

FICHTER, Darlene. Administrative and Factual Metadata for Intranets: Issues and Options. *Online*, 1999, vol. 3, nº 6, p. 88-90.

FITZWATER, Larry and Linda Spencer. EPA Terminology Reference System (TRS) [presentación PPT]. En: *CENDI Conference Controlled Vocabulary and The Internet (1999. Bethesda)*. CENDI, Subject Analysis and Retrieval Working Group, 1999. Disponible en: <http://www.dtic.mil/cendi/presentations/Spencer-Fitzawater.ppt>

Bibliografía

FLOYD, Michael. *Creación de sitios Web con XML*. Madrid: Pearson Education, 2000, 364 p.

FRENCH, James C., Allison L. Powell, Walter R. Crighton, III. Efficient Searching in Distributed Digital Libraries. En: *ACM Digital Libraries (3. 1998. Pittsburg)*. Ian Witten, Rob Akscyn, Frank M. Shipman, III, eds. New York: ACM, 1998, p. 283-284.

FREESE, Eric *TopicMap vs. RDF* [documento HTML]. London: Open University, 2000. Disponible en: <http://ep.open.ac.uk/PubSys/resources/html/free0000.html> (consultado el 17 de julio de 2001).

FRETER, Todd. *XML: Mastering Information on the Web* [documento HTML]. Sun Microsystems, 10 de marzo de 1998. Disponible en: <http://www.sun.com/980310/xml>

FREW, James, et al. Generic Query Metadata for Geospatial Digital Libraries [documento HTML]. En: *IEEE Meta-data Conference (3. 1999. Bethesda)*. Computer Society, rev. 9 de febrero de 1999. Disponible en: <http://computer.org/proceedings/meta/1999/papers/55/jfrew.htm>

FRÍAS, José Antonio. La catalogación de los documentos electrónicos: problemática, propuestas e iniciativas. En: *Sistemas de Información. Balance de 12 años de jornadas. Jornadas españolas de documentación automatizada (5. 1996. Cáceres)*. Cáceres: Universidad de Extremadura, Servicio de Publicaciones; ABADMEX, 1996, vol. 1, p. 431-438.

FURNER, Jonathan. IR on the Web: An Overview. *Vine (Theme issue: Digital Images in Libraries)*, July 1998, nº 107, p. 3-13.

G

GALE Adopts Sfx Open Url Standard [documento HTML]. Galegroup, Thompson Learnig, 31 de octubre de 2001. Disponible en: http://www.galegroup.com/press_room/article/oct01_sfx.htm

GALLART MARSILLAS, Núria. La biblioteca digital distribuïda: situació del projecte DecomateII. En: *Jornades Catalanes de Documentació (7. 1999. Barcelona)*. Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 1999, p. 399-404. También disponible en Internet en: <http://www.cobdc.org/7es/17.pdf>

GARCÍA GÓMEZ, Juan Carlos, Francisco Javier Martínez Méndez, José Antonio Gómez Hernández. Bibliotecas digitales de Internet como servicio complementario para pequeñas bibliotecas. En: *Jornades Catalanes de Documentació (6. 1997. Barcelona)*. Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 1997, p. 275-284. También disponible en Internet en: <http://www.um.es/gtiweb/fjmm/bibdigital.htm>

Bibliografía

GARCÍA MARTÍNEZ, Ana María. Definición y estilo de los objetos de información digitales y metadatos para la descripción. *Boletín de la Asociación Andaluza de Bibliotecarios*, 2001, vol. 16, nº 63, p. 23-47.

GARRET, John. Digital Libraries: The Grand Challenges [documento TXT]. *EDUCOM Review*, July/August 1993, vol. 28, nº 4,. Disponible en: <http://www.ifla.org/documents/libraries/net/garret.txt> (consultado el 16 de septiembre de 1999).

GAYNOR, Edward. *From MARC to Markup: SGML and Online Library Systems* [documento HTML]. Charlottesville: University of Virginia Library, Special Collections, 8 de mayo de 1996. Disponible en: http://www.lib.virginia.edu/speccol/scdc/articles/alcts_brief.html

GETTY INFORMATION INSTITUTE. *Standards, Metadata and Vocabulary Tools for Cultural Heritage Documentation and Access: An Overview*. [presentación PPT]. Getty Information Institute, 4 de junio de 1998, 64 slides, (ejemplar cedido por Murtha Baca, cooaut.).

GILL, Tony. The Dublin Core Metadata Element Set: Useful Tool or the Emperor's Newest Clothes? *Spectra*, Winter 1997, p. 40-46.

GINGER, Katy *DLESE Search Buckets: Details Search Buckets Concepts and Schemes for Interoperability* [documento HTML]. DLESE.org, rev. 4 de diciembre de 2000. Disponible en: <http://www.dlese.org/Metadata/buckets.htm> (consultado el 4 de junio de 2001).

GLASSEL, Aimee. Was Ranganathan a Yahoo? [documento HTML]. *InterNIC News*, 1998, nº 3. Disponible en: <http://scout.cs.wisc.edu/addserv/toolkit/enduser/archive/1998/euc-9803.html>

GLOSSARY [documento HTML]. En: *Schemas: Forum for Metadata Schema Implementers*. Maintained by Manjula Patel. Schemas, 2 de agosto de 2000. Disponible en: <http://www.schemas-forum.org/related/glossary.html>

GOLDMAN, Roy, Jason McHugh and Jennifer Widom. From Semistructured Data to XML. *Markup Languages: Theory and Practice*, Spring 2000, vol. 2, issue 2, p. 153-163.

GÓMEZ HERNÁNDEZ, José Antonio. Legitimación y funciones de la biblioteca en el contexto de la sociedad digital. *Scire*, 1998, vol. 4, nº. 2, p. 63-77.

GÓMEZ SKÁRMETA, Antonio, Eduardo Martínez Graciá, Angélica M^a Galiano Romero. Una arquitectura de metadatos para la gestión de la información en el web. *Novática*, julio/agosto 2000, nº 146, p. 8-11.

GONÇALVES, Marcos André, et al. MARIAN Searching and Querying Across Heterogeneous Federated Digital Libraries [documento PDF]. En: *DELOS Network*

Bibliografía

of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries (2000. Zurich). ERCIM Workshop Proceedings, N° 01/W001 [2001]. Disponible en: http://www.ercim.org/publication/ws-proceedings/DelNoe01/11_Fox.pdf

GONÇALVES, Marcos André, Claudia Bauzer Medeiros. Constructing Geographic Digital Libraries Using a Hypermedia Framework. *Multimedia Tools and Applications*, 1999, n° 8, p. 341-356.

GOODCHILD, Michael F. *Alexandria Digital Library: Report on Workshop on Metadata* [documento HTML]. Santa Bárbara: University of California, 8 de noviembre de 1995. Disponible en: http://www.alexandria.ucsb.edu/public-documents/metadata/etaddata_ws.html

GORMAN, Michael. From Card Catalogues to WebPACS [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 1 de noviembre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/gorman_paper.html (consultado el 21 de febrero de 2001).

GORMAN, Michael. What is the Future of Cataloguing and Cataloguers? [documento HTML]. En: *IFLA General Conference (63. 1997. Copenhagen)*. Ottawa: Biblioteca Nacional de Canadá, IFLA, 1997. Disponible en: <http://www.nlc-bnc.ca/ifla/IV/ifla63/63gorm.htm> (consultado el 20 de agosto de 1998).

GRADMANN, Stefan. Cataloguing vs. Metadata: Old Wine in New Bottles? [documento HTML]. En: *IFLA General Conference (64. 1998. Amsterdam)*. París: IFLA; INIST, rev. 23 de junio de 1998. Disponible en: <http://ifla.inist.fr/ifla/IV/ifla64/007-126e.htm> Publicado posteriormente en: *International Cataloguing and Bibliographic Control*, Oct./Dec. 1999, vol. 28, n° 4, p. 88-90.

GRAHAM, Stephen. Instances of Availability – the HeadLine Resource Data Model. *Vine (Theme issue: Metadata. Part 2)*, 1999, n° 117, p. 13-17.

GRANGER, Stewart. Metadata and Digital Preservation: A Plea for Cross-interest Collaboration. *Vine (Theme issue: Metadata. Part 2)*, 1999, n° 117, p. 24-29.

GRASSIAN, Esther. *Thinking Critically about World Wide Web Resources* [documento HTML]. California: UCLA College Library, junio de 1995, rev. 6 de septiembre de 2000. Disponible en: <http://www.library.ucla.edu/libraries/college/help/critical/index.htm>

GREEN, David. Beyond Word and Image. Networking Moving Images: More than Just the “Movies” [documento HTML]. *D-Lib Magazine*, July/August 1997, vol. 3, n° 7/8. Disponible en: <http://www.dlib.org/dlib/july97/07green.html>

GREEN, David. The Evolution of Web Searching. *Online Information Review*, 2000, vol. 24, n° 2, p. 124-137.

Bibliografía

GREENBERG, Jane. A Comparison of Web Resource Access Experiments: Planning for the New Millennium [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 29 de septiembre de 2000, 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/greenberg_paper.html

GREISDORF, Howard. Relevance: an Interdisciplinary and Information Science Perspective. *Informing Science*, 2000, vol. 3, nº 2, p. 67-72. También disponible en Internet en: <http://inform.nu/Articles/Vol3/v3n2p67-72.pdf>

GRIFONI, Giovanna. Come orientarsi tra i motori di ricerca. Una panoramica sugli strumenti di recupero delle informazioni in Internet. *Biblioteche oggi*, giugno 1997, XV, nº 5, p. 10-16. También disponible en: <http://www.burioni.it/forum/grif-mot.htm>

GRÖTSCHEL, Martin, Joachim Lügger. Scientific Information Systems and Metadata [documento HTML]. *ZIB Report SC-98-26*. Berlin: Konrad-Zuse-Zentrum für Informationstechnik, 26 de octubre 1998. Disponible en: <http://elib.zib.de/ftp/pub/UserHome/Luegger/Dresden/Metadata.htm>

GROUT, Catherine, Phill Purdy, Janine Rymer. *Creating Digital Resources for the Visual Arts: Standards and Good Practice* [documento HTML]. Bristol: AHDS, VADS, rev. 9 de abril de 2001. Disponible en: http://vads.ahds.ac.uk/guides/creating_guide/contents.html

GROUT, Catherine and Tony Gill. *Visual Arts, Museum & Cultural Heritage Metadata draft workshop report* [documento HTML]. Visual Arts Data Service, rev. 6 de agosto de 1997. Disponible en: http://vads.ahds.ac.uk/training_advice/standards/Metadata_workshop/Metadata_index.html (consultado el 2 de abril de 2001).

GRUBER, Tom. *What is an Ontology* [documento HTML]. Stanford: Stanford University, Knowledge Systems Laboratory, rev. 12 de septiembre de 1997. Disponible en: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>

GUENTHER, Rebecca. Naming Conventions for Digital Resources [documento HTML]. En: *MARC Homepage*. Washington, DC.: Library of Congress, 8 de julio de 1998. Disponible en: <http://lcweb.loc.gov/marc/naming.html>

GUENTHER, Rebecca S. Re: RFCs and NISO Standardization of the DC. [correo electrónico en lista de distribución]. *META2*, 13 de enero de 1998; 13:56. Disponible en: <http://www.mailbase.ac.uk/lists/dc-general/1998-01/0013.html>

A GUIDE to the ANSI/NISO Z39.50 Protocol. Bethesda: National Information Standards Organization, 1995.

GUIDELINES for Coding Electronic Resources in Leader/06 [documento HTML]. Washington, D. C.: Library of Congress, Network Development and MARC

Bibliografía

Standards Office, rev. 7 de enero de 1999. Disponible en: <http://lcweb.loc.gov/marc/ldr06guide.html>

GUIDELINES for Markup of Electronic Text [documento HTML]. Peter C. Gorman, ed. Madison: University of Wisconsin Library, 11 de octubre de 2000. Disponible en: <http://www.library.wisc.edu/help/tech/TEI/guidelines.html>

GUIDELINES for the Use of Field 856 [documento HTML]. Washington, D. C.: Library of Congress, Network Development and MARC Standards Office, rev. 6 de noviembre de 2001. Disponible en: <http://www.loc.gov/marc/856guide.html>

GUIDELINES on Best Practices for Using Electronic Information: How to Deal with Machine-readable Data and Electronic Documents. Updated and enlarged ed. Luxembourg: Office for Official Publications of the European Communities, 1997. También disponible en Internet en: <http://europa.eu.int/ISPO/dlm/documents/gdlines.pdf>

GUINCHARD, Carolyn. Summary of DC-Libraries Questionnaire Responses [documento RTF]. En: *Survey Results: Dublin Core Use in Libraries* [correo electrónico en lista de distribución]. *DC-LIBRARIES*, Miércoles 25 de abril de 2001; 16:03:45. Disponible en: <http://www.jiscmail.ac.uk/cgi-bin/wa.exe?A2=ind0104&L=dc-general&F=&S=&P=2403>

H

HAAS, Stephanie W., Erika S. Grams. Page and Link Classifications: Connecting Diverse Resources. En: *ACM Digital Libraries (3. 1998. Pittsburg)*. Ian Witten, Rob Akscyn, Frank M. Shipman, III, eds. New York: ACM, 1998, p. 99-107.

HAKALA, Juha. Internet Metadata and Library Cataloguing. *International Cataloguing and Bibliographic Control*, January/March 1999, vol. 28, n° 1, p. 21-25.

HAKALA, Juha, et al. *The Nordic Metadata Project* [documento HTML]. Helsinki University Library, 25 de junio de 1998. Disponible en: <http://www.lib.helsinki.fi/meta/nmfinal.htm>

HAKALA, Juha, Ole Husby and Traugott Koch. *Warwick Framework and Dublin Core Set Provide a Comprehensive Infrastructure for Network Resources: Report from the Metadata Workshop II, Warwick, UK, April 1-3, 1996* [documento HTML]. Lunds University Electronic Library, 6 de abril de 1996, rev. 19 de junio de 1996. Disponible en: <http://www.ub2.lu.se/tk/warwick.html>

HANE, Paula J. Beyond Keyword Searching: Oingo and Simpli.com Introduce Meaning-Based Searching [documento HTML]. *Information Today*, 20 December 1999. Disponible en: <http://www.infotoday.com/newsbreaks/nb1220-2.htm> (consultado el 23 de octubre de 2000).

Bibliografía

HARDY, Darren R., Michael F. Schwartz and Duane Wessels. *Harvest User's Manual (Version 1.4 patchlevel 2)* [documento HTML]. Harvest, 31 de enero de 1996, rev. 22 de abril de 1996. Disponible en: <http://harvest.transarc.com/afs/transarc.com/public/trg/Harvest/user-manual/user-anual.html> (consultado el 18 de junio de 1998); Disponible en: <http://www.tardis.ed.ac.uk/harvest/docs/old-manual/user-manual.html>

HARMSSEN, Bernd. Adding Value to Web-OPACs. *The Electronic Library*, 2000, vol. 18, n° 2, p. 109-113.

HARTER, Stephen P. Scholarly Communication and the Digital Library: Problems and Issues [documento HTML]. *Journal of Digital Information*, 1997, vol. 1, issue 1. Disponible en: <http://jodi.ecs.soton.ac.uk/Articles/v01/i01/Harter>

HARVEY, Francis. Quality Means More than Standards: Development of an Interface for Better Use of Metadata. *GIM International*, June 2000, vol. 14, n° 6, p. 77-80.

HEANEY, Michael. Object-Oriented Cataloging. *Information Technology and Libraries*, September 1995, p. 135-153.

HEERY, Rachel. *DI.1 Metadata Formats* [documento RTF]. En: *WP Study of Metadata*, issue 1.0. Bath: University, UKOLN, 23 de diciembre de 1996, 112 p. Disponible en: <http://hosted.ukoln.ac.uk/biblink/wp1/d1.1.rtf>

HEERY, Rachel. *Review of Metadata Formats* [documento HTML]. Bath: University, UKOLN, maintained by Michael Day, rev. 28 de enero de 1998. Disponible en: <http://www.ukoln.ac.uk/metadata/review.html> Pre-publicación de: Rachel Heery. Review of Metadata Formats. *Program*, October 1996, vol. 30, n° 4, p. 345-373.

HEERY, Rachel, Andy Powel and Michael Day. Metadata: CrossROADS and Interoperability [documento HTML]. *Ariadne*, 11-March-1998, issue 14. Disponible en: <http://www.ariadne.ac.uk/issue14/metadat> (consultado el 25 de junio de 1998).

HEERY, Rachel and Manjula Patel. Application Profiles: Mixing and Matching Metadata Schemas [documento HTML]. *Ariadne*, 9-October-2000, issue 25. Disponible en: <http://www.ariadne.ac.uk/issue25/app-profiles>

HIGGINS, Margaret. Meta-Information and Time: Factors in Human Decision Making. *Journal of the American Society for Information Science*, 1999, vol. 50, n° 2, p. 132-139.

HILL, Linda. Re:NKOS: Who Assign Metadata? [correo electrónico en lista de distribución] *NKOS*, 29 de julio de 1998; 17:02 [no existe archivo de la lista consultable de esa fecha]. Disponible en emendez@bib.uc3m.es

HILL, Linda L., et al. Collection Metadata Solutions for Digital Library Applications. *Journal of the American Society for Information Science (Special Topic Issue:*

Bibliografía

Integrating Multiple Overlapping Metadata Standards), 1999, vol. 50, nº 13, p. 1169-1181.

HÍPOLA, Pedro, Benjamín Vargas-Quesada y José A. Senso. Bibliotecas digitales: situación actual y problemas. *El profesional de la información*, abril 2000, vol. 9, nº 4, p. 4-13.

HÍPOLA, Pedro, Ricardo Eito Brun. Edición digital: formatos y alternativas. *El profesional de la información*, octubre 2000, vol. 9, nº 10, p. 4-15.

HISOFTWARE TagGen™ Office: User's Guide [documento PDF, comprimido ZIP]. Concord: Hiawatha Island Software Company, rev. 21 de junio de 2001. Disponible en: <http://www.hisoftware.com/DocCentral/taggenpdf.zip>

HJELM, Johan. *Creating the Semantic Web with RDF: Professional Developer's Guide*. New York, etc.: Wiley Computer Publishing, John Wiley & Sons, 2001.

HJØRLAND, Birger. Towards a Theory of Aboutness, Subject, Topicality, Theme, Domain, Field, Content... and Relevance. *Journal of the American Society for Information Science and Technology*, 2001, vol. 59, nº 9, p. 774-778.

HODGE, Gail. *System of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. Washington: Council on Library and Information Resources, April 2000.

HODGSON, Katrina. Metadata: Foundations, Potential and Applications [documento HTML]. Edmonton: University of Alberta, School of Library & Information Studies, 31 de marzo de 1998, rev. 29 de abril de 1998. Disponible en: <http://www.slis.ualberta.ca/538/khodgson/metadata.htm>

HORN, François. Diversité des informations traitées par des moyens informatiques, standardisation optimale et acteurs su processus de standardisation [documento HTML]. *Solaris*, Decembre 1999/Janvier 2000. Disponible en: <http://www.info.unicaen.fr/bnum/jelec/Solaris/d06/6horn.html>

HOWE, Denis. *FOLDDOC: Free Online Dictionary of Computing* [documento HTML]. Foldoc 1993, rev. 22 de septiembre de 2001. Disponible en: <http://wombat.doc.ic.ac.uk>

HPCC FY 1995 Implementation Plan [documento HTML]. National Coordination Office for Computing, Information and Communications, 26 de mayo de 1994. Disponible en: <http://www.ccic.gov/pubs/imp95>

HSIEH-YEE, Ingrid. Modifying Cataloging Practice and OCLC Infrastructure for Effective Organization of Internet Resources [documento HTML]. En: *OCLC Internet Cataloging Colloquium (1996. San Antonio)*. OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/hsieh.htm>

Bibliografía

HUBER, Alexander. Anglistik Guide and History Guide at the State and University Library at Goettingen (SUB Goettingen): Internet-based subject gateways on British and American literature and history. En: *Digital resources for the humanities: Confernce abstracts*, 2001, p. 47-50.

HUC, Claude, Thierry Levoir and Michel Nonon-Latapie. Metadata: Models and Conceptual Limits [documento HTML]. En: *IEEE Metadata Conference (2. 1997. Silver Spring)*. IEEE Computer Society, 1997. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/chuc/chuc.html>

HUDGINS, Jean, Grace Agnew, Elizabeth Brown. *Getting Mileage out of Metadata: Applications for the Library*. Chicago: ALA, 1999.

HUDGINS-BONAFIELD, Christine. Who will Master Metadata? [documento HTML]. *Network Computing*, 2000. Disponible en: <http://www.networkcomputing.com/608/608business.html>

HUNTER, Jane. MetaNet: A Metadata Term Thesaurus to Enable Semantic Interoperability Between Metadata Domains [documento HTML]. *Journal of Digital Information*, 2001, vol. 1, issue 8. Disponible en: <http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter>

HUNTER, Jane, Liz Armstrong. A Comparison of Schemas for Video Metadata Representation [documento HTML]. En: *International World Wide Web Conference (8. 1999. Toronto)*, rev. 5 de diciembre de 2000. Disponible en: <http://www8.org/w8-papers/3c-hypermedia-video/comparison/comparison.html>

HUNTER, Jane, Carl Lagonze. Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles [documento HTML]. Brisbane: University of Queensland, Distributed Systems Technologies, rev. 22 de abril de 2001. Disponible en: <http://archive.dstc.edu.au/RDU/staff/jane-hunter/www10/paper.html>

HUTHWAITE, Ann. AACR2 and its Place in the Digital World: Near-Term Solutions and Long-Term Direction [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 30 de junio de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/huthwaite_paper.html

HUWE, Terence K. New Search Tools for Multidisciplinary Digital Libraries. *Online*, March 1999, vol. 23, n° 2, p. 67-70. También disponible en Internet en: <http://www.onlineinc.com/onlinemag/OL1999/huwe3.html>

HUXLEY, Lesly. Renardus: Fostering Collaboration Between Academic Subject Gateways in Europe. *Online Information Review*, 2001, vol. 25, n° 2, p. 121-127.

Bibliografía

HYNES, Chris. *Re: METASTAR* [correo electrónico personal], 14 de mayo de 2001; 14:34. Disponible en: emendez@bib.uc3m.es

HYPERION: sistema de archivos de documentos digitales [presentación PPT]. En: *Jornadas de Tecnología de la Información: Sistemas de Gestión Integrada de Bibliotecas* (2000. Bahía Blanca) [CD-ROM]. Bahía Blanca: Universidad Nacional del Sur, Biblioteca Central, 2001.

I

IANNELLA, Renato. *Mostly Metadata: a Bit Smarter Technology* [documento HTML]. En: *Resource Discovery Unit*. Brisbane: University of Queensland, Distributed Systems Technologies, 1997. Disponible: <http://www.dstc.edu.au/RDU/reports/VALA1998>

IANNELLA, Renato. *Resource Discovery Project* [documento HTML]. *Ariadne*, 7-March-1997, issue 8. Disponible en: <http://www.ariadne.ac.uk/issue8/resource-discovery>

IANNELLA, Renato, Andrew Waugh. *Metadata: Enabling the Internet* [documento HTML]. En: *Resource Discovery Unit*. Brisbane: University of Queensland, Distributed Systems Technologies, rev. 4 de febrero de 1997. Disponible: <http://www.dstc.edu.au/RDU/reports/CAUSE97>

IOANNIDES, Demetrios. *XML an Electronic Publishing: XML Schema Languages: Beyond DTD*. *Library HiTech*, 2000, vol. 18, n° 1, p. 9-14.

IFLA. *Digital Libraries: Metadata Resources* [documento HTML]. The Hague: IFLA, 1995-2000, rev. 23 de febrero de 2001. Disponible en: <http://www.ifla.org/II/metadata.htm> y en: <http://ifla.inist.fr/II/metadata.htm>

IFLA. *ISBD (ER) International Standard Bibliographic Description for Electronic Resources: revised from the ISBD (CF) International Standard Bibliographic Description for Computer Files*. München: Saur, 1997. También disponible en Internet en: <http://www.ifla.org/VII/s13/pubs/isbd.htm>

INFORMATION retrieval: data structures & algorithms. William B. Frakes, Ricardo Baeza-Yates, eds. New Jersey: Prentice Hall, 1992.

INTERNET, metadatos y acceso a la información en bibliotecas y redes en la Era Electrónica. Filiberto Felipe Martínez Arellano, Lina Escalona Ríos, comp. México: UNAM, Centro Universitario de Investigaciones Bibliotecológicas; Infoconsultores, 2000.

INTERNET Searching and Indexing: The Subject Approach. Alan R. Thomas, James R. Shearer, eds. New York, London, Oxford: The Haworth Press, 2000.

Bibliografía

INTRODUCCIÓN a los Metadatos: vías a la información digital. Murtha Baca, ed. Los Angeles: J. Paul Getty Trust, 1999.

INTRODUCTION to Metadata: Pathways to Digital Information. [version 2.0] [documento HTML]. Murtha Baca, ed. Los Angeles: Getty Research Institute, 1 de agosto 2000, rev. 13 de septiembre de 2001. Disponible en: <http://www.getty.edu/research/institute/standards/intrometadata/index.html>

INTRODUCTION to Vocabularies: Enhancing Access to Cultural Heritage Information. Los Angeles: J. Paul Getty Trust, 1998.

IP, Albert, et al. Metasearching or Megasearching: Toward a Data Model for Distributed Resource Discovery. En: *e-Education: challenges and opportunities. Hong Kong Web Symposium (5. 1999. Hong Kong)*. Fred Castro, Robert Lai and Sr. Margaret Wong, eds. Hong Kong: Social Sciences Research Centre, 1999, p. 65-82. También disponible en Internet en: <http://www.dls.au.com/metadata/DataModel.html>

ISO/IEC 15445:2000 / DCOR 1:2001(E). *Information Technology - Document Description and Processing Languages - HyperText Markup Language (HTML)* [documento HTML]. Dublin: Trinity College, Computer Science Department, 15 de mayo de 2000. Disponible en: <http://www.cs.tcd.ie/15445/15445.html>

ISO 8824 — *Information Processing Systems - Open Systems Interconnection - Specification of Abstract Syntax Notation One (ASN.1)*, 1990.

J

JACKSON, Joe and Donald L. Gilstrap. XML and Better Web Searching. *Library Hi Tech*, 1999, vol. 17, nº 3, p. 316-320.

JEFFERY, Keith G. Metadata: An Overview and Some Issues [documento PDF]. En: *ERCIM Database Research Group Workshop on Metadata for Web Databases (11. 1998. Sankt Augustin)*. ERCIM Workshop Proceedings, Nº 98/W002, 1998. Disponible en: <http://www.ercim.org/publication/ws-proceedings/11th-EDRG/jefferey.pdf>

JENKINS, Charlotte, et al. Automatic RDF Metadata Generation for Resource Discovery [documento HTML]. En: *International World Wide Web Conference (8. 1999. Toronto)*, rev. 5 de diciembre de 2000. Disponible en: <http://www8.org/w8-papers/2c-search-discover/automatic/automatic.html>

JENSEN, Bob. *Jensen's Technology Glossary* [documento HTML]. San Antonio: Trinity University, rev. 21 de noviembre de 2001. Disponible en: <http://www.trinity.edu/~rjensen/245glosf.htm>

Bibliografía

JIMÉNEZ GUARDIOLA, Ivette Patricia y Carlos Arturo Parra Llanos. *El uso de RDF en el diseño de sistemas de información en la web y su aplicación en el prototipo Oceandata*. 2000, 43 p., (ejemplar cedido por Carlos Parra, coaut.).

JOINT Funding Council's Libraries Review Group: Report (The Follet Report) [documento HTML]. Bath: University, UKOLN, December 1993, rev. 19 de febrero de 1999. Disponible en: <http://www.ukoln.ac.uk/services/papers/follett/report>

JOINT INFORMATION SYSTEMS COMMITTEE (JISC). *eLib: Electronic Library Program* [documento HTML]. Bath: University, UKOLN, rev. 18 de octubre de 2001. Disponible en: <http://www.ukoln.ac.uk/services/elib>

JUL, Eric. Cataloging Internet Resources: Survey and Prospectus. *Bulletin of the American Society for Information Science*, October/November 1997, vol. 24, n° 1, p. 6-9.

K

KAHN, Robert and Robert Wilensky. *A Framework for Distributed Digital Object Services* [documento HTML]. París: IFLA, 13 de mayo de 1995, rev. 23 de junio de 1998. Disponible en: <http://ifla.inist.fr/ifla/documents/libraries/cataloging/metadata/k-w.htm> (consultado el 9 de julio de 1998).

KASHYAP, Virpul, Amit Sheth. *Information Brokering Across Heterogeneous Digital Data: A Metadata-based Approach*. Boston, Dordrecht, London: Kluwer Academic Publishers, 2000.

KASHYAP, Virpul, Amit Sheth. *Semantic Heterogeneity in Global Information Systems: the Role of Metadata, Context and Ontologies* [documento PS]. Georgia: University of Georgia. Department on Computer Sciences. Large Scale Distributed Information Systems, 12 de junio de 1996. Disponible en: <http://lsdis.cs.uga.edu/lib/download/KS97.ps> (consultado el 12 de febrero de 2001).

KASPAROVA, N & M. Shwartsman. Creación de la base de metadatos de recursos electrónicos en Rusia: problemas y perspectivas. *IFLA Council and General Conference (66. 2000. Jerusalem.)* [documento HTML]. IFLA, 18 de julio de 2000. Disponible en: <http://www.ifla.org/IV/ifla66/papers/139-168s.htm>

KEEN, Michael. Using Channels, Topics, Keywords and Other Metadata in an Electronic Alerting Service. *Vine (Theme issue: Metadata. Part 2)*, 1999, n° 117, p. 3-12.

KELLY, Brian. Web Focus: Using the Web to Promote Your Web Site [documento HTML]. *Ariadne*, 8-Dec-1999, issue 22. Disponible en: <http://www.ariadne.ac.uk/issue22/web-focus>

Bibliografía

- KENA, Jenny. The Role and Development of Metadata [documento HTML]. *Ozemail*, september 1997. Disponible en: <http://www.ozemail.com.au/~jkena/metadata.html>
- KETCHELL, John. Standardization in the Information Society – or How to Avoid Market Anarchy [documento HTML]. En: *European Conference on Research and Advanced Technology for Digital Libraries (4. 2000. Lisbon): ECDL2000*. Disponible en: <http://www.ibl.pt/org/agenda/ecdl2000/ketchell.html> (consultado el 25 de abril de 2001).
- KIELY, Don. XHTML: A Bridge to the Future [documento HTML]. *Information Week Online*, 8 de mayo de 2000. Disponible en: <http://www.informationweek.com/785/language.htm>
- KIM, Hyun-Hee and Chang-Seok Choi. XML: How it Will Be Applied to Digital Library Systems. *The Electronic Library*, 2000, vol. 18, nº 3, p. 183-189.
- KIRRIEMUIR, John, et al. Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach [documento HTML]. *D-Lib Magazine*, January 1998, vol. 4, nº 1. Disponible en: <http://www.dlib.org/dlib/january98/01kirriemuir.html>
- KNIGHT, Jon. Will Dublin form the Apple Core?[documento HTML]. *Ariadne*, 23-January-1997, issue 7. Disponible en: <http://www.ariadne.ac.uk/issue7/mcf/intro.html>
- KLAUS, Helmut. Understanding of Thesauri in the Online Dialogue: Implications for Construction and Display. *Cataloguing Australia*, 1999, vol. 25, nº 1/4, p. 120-139.
- KOCH, Traugott *Browsing and Searching Internet Resources* [documento HTML]. Lund University Electronic Library, junio de 1994, rev. 19 de julio de 2000. Disponible en: http://www.ub2.lu.se/nav_menu.html
- KOCH, Traugott. *Controlled Vocabularies, Thesauri and Classification Systems Available in the WWW. DC Subject* [documento HTML]. Lund: University Libraries, rev. 27 de septiembre de 2001. Disponible en: <http://www.lub.lu.se/metadata/subject-help.html>
- KOCH, Traugott. *Documentation to: Metadata Indexing and Searching in Large Search Services*. [documento HTML]. Lund University Electronic Library, 8 de enero de 1998, rev. 26 de febrero de 1998. Disponible en: <http://www.lub.lu.se/tk/metadata/MDsearch.html>
- KOCH, Traugott. *Searching the Web. Systematic Overview Over Indexes Services* [documento HTML]. Lund University Electronic Library, febrero de 1995, rev. 25 de febrero de 1998. Disponible en: http://www.ub2.lu.se/tk/websearch_systemat.html
- KOEHLER, Wallace. Digital Libraries and World Wide Web Sites Page Persistence [documento HTML]. *Information Research*, June 1999, vol. 4, nº 4. Disponible en: <http://www.shef.ac.uk/~is/publications/infres/paper60.html>

Bibliografía

KOMITO, Lee. Electronic Communities in an Information Society: Paradise. Mirage, or Malaise? *Journal of Documentation*, January 2001, vol. 57, n° 1, p. 115-129.

KOSTER, Martijn. The Web Robots FAQ... [documento HTML] The Web Robots Pages, rev. 10 de julio de 2001. Disponible en: <http://www.robotstxt.org/wc/faq.html>

KOVACS, Diane K. And Angela Elkordy. Collection Development in Cyberspace: Building an Electronic Library Collection. *Library Hi Tech*, 2000, vol. 18, n° 4, p. 335-359.

KUMAR, Vijay, Richard Furuta, Robert B. Allen. Metadata Visualization for Digital Libraries: Interactive Timeline Editing and Review. En: *ACM Digital Libraries (3. 1998. Pittsburg)*. Ian Witten, Rob Akscyn, Frank M. Shipman, III, eds. New York: ACM, 1998, p. 126-133.

KUNY, Terry and Gary Cleveland. The Digital Library: Myths and Challenges. *IFLA Journal*, 1998, vol. 24, n° 2, p. 107-113. También disponible en Internet en: <http://www.ifla.org/IV/ifla62/62-kuny.pdf>

KUNZE, J. *Encoding Dublin Core Metadata in HTML: RFC 2731* [documento TXT]. IETF, Network Working Group, December 1999. Disponible en: <http://www.ietf.org/rfc/rfc2731.txt>

L

LAFON, Yves, Bert Bos. Describing and Retrieving Photos Using RDF and http: W3C Note [documento HTML]. W3C, 28 de septiembre de 2000. Disponible en: <http://www.w3.org/TR/photo-rdf>

LAGOZE, Carl. Business unusual: How "Event-Awareness May Breathe Life into Catalog? [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 19 de octubre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/lagonze_paper.html (consultado el 21 de febrero de 2001).

LAGOZE, Carl. The Warwick Framework: A Container Architecture for Diverse Sets of Metadata [documento HTML]. *D-Lib Magazine*, July/August 1996, vol. 2, n° 7/8. Disponible en: <http://www.dlib.org/dlib/july96/lagonze/07lagonze.html> (consultado el 19 de agosto de 1999).

LAGOZE, Carl, et al. Using Web Metadata: part 2, the Resource Description Framework [presentación PPT]. *WWW7 Tutorial*. OCLC, 10 de abril de 1998. Disponible en: <http://purl.oclc.org/~emiller/talks/www7/tutorial/part2>

Bibliografía

- LAGOZE, Carl, David Fielding. Defining Collections in Distributed Digital Libraries [documento HTML]. *D-Lib Magazine*, November 1998, vol. 4, n° 11. Disponible en: <http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>
- LAGONZE, Carl, Herbert van de Sompel. The Open Archives Initiative: Building a Low-Barrier Interoperability Framework. En: *ACM/IEEE-CS Joint Conference on Digital Libraries (2001. Raonoke)*. New York: ACM, 2001, p. 54-62.
- LAM, K.T. *Moving from MARC to XML* [documento HTML]. Hong Kong University of Science and Technology, rev. 24 de mayo de 2001. Disponible en: <http://ihome.ust.hk/~lblkt/xml/marc2xml.html>
- LANCASTER, Frederick W. *El control del vocabulario en la recuperación de información*. Valencia: Universitat de Valencia, 1995.
- LANDOW, George P. *Hipertexto: la convergencia de la teoría crítica contemporánea y la tecnología*. Barcelona: Paidós, 1995.
- LARGE, Andrew, Lucy A. Tedd and R. J. Hartley. *Information Seeking in the Online Age: Principles and Practice*. London, etc.: Bowker Saur, 1998.
- LARSON, Carolyn, Linda Arret. Descriptive Resource Needs from the Reference Perspective [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 26 de octubre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/arret_paper.html
- LARSON, Ray R. Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace. En: *ASIS Annual Meeting (59. 1996. Baltimore)*. Medford: Information Today, 1996, p. 71-78.
- LARSON, Ray R. Distributed Resource Discovery: Using Z39.50 to Build Cross-Domain Information Servers. En: *Joint Conference on Digital Libraries (2001. Roanoke)*. New York: ACM, 2001, p.52-53.
- LASSILA, Ora. *Introduction to RDF Metadata* [documento HTML]. World-Wide Web Consortium, 13 de noviembre de 1997. Disponible en: <http://www.w3.org/TR/NOTE-rdf-simple-intro>
- LASSILA, Ora. Web Metadata: A Matter of Semantics. *IEEE Internet Computing*, July-August 1998, p.30-37.
- LAWRENCE, Steve, Giles, C. Lee. Accessibility of Information on the Web. *Nature*, 8 July 1999, vol. 400, p. 107-109.

Bibliografía

LAWRENCE, Steve, Giles, C. Lee. Searching the World Wide Web. *Science*, 3 April 1998, vol. 280, n° 5360, p. 98-100. También disponible en Internet en: <http://www.neci.nj.nec.com/~lawrence/science98.html>

LE VAN, Ralph. *Dublin Core and Z39.50* [documento XHTML]. DCMI, 2 de febrero de 1998. Disponible en: <http://www.dublincore.org/documents/1998/02/02/dc-z3950>

LE VAN, Ralph. *Dublin Core and Z39.50: Draft Version 1.2* [documento HTML]. OCLC, Office and Research Special Projects, 2 de febrero de 1998. Disponible en: <http://cypress.dev.oclc.org:12345/~rrl/docs/dublincoreandz3950.html>

LEASE MORGAN, Eric. Comparing Open Source Indexers [documento HTML]. *Infomotions Musings*, May 29, 2001. Disponible en: <http://www.infomotions.com/musings/opensource-indexers>

LELOUP, Catherine. *Motores de búsqueda e indexación: entornos cliente servidor, Internet e Intranet*. Barcelona: Ediciones Gestión 2000, 1998.

LEVINE Marilyn M. A Brief History on Information Brokering [documento HTML]. *Bulletin ASIS*, February/March 1995, vol. 21, n° 3. Disponible en: <http://www.asis.org/Bulletin/Feb-95/levine.html>

LIBRARY of Congress Digital Repository Development: Core Metadata Elements [documento HTML]. Washington, DC.: Library of Congress, 18 de mayo de 2000. Disponible en: <http://lcweb.loc.gov/standards/metadata.html>

LINDSTROM, Richard W., Mary S. Woodley. Irresistible Metadata: Guidelines for Usage of Dublin Core Metadata in Online Exhibitions. *Spectra*, Spring 1999, vol. 26, issue 1, p. 19-31.

LLORET ROMERO, Nuria. Cómo plantear un proyecto para el desarrollo de una biblioteca digital. En: *Jornadas Andaluzas de Documentación (2. 1999. Granada)*. Granada: Asociación Andaluza de Documentalistas, 1999, p. 307-317.

LÓPEZ, Diego R. Webber: Uso de componentes para la armonización de contenidos y metadatos. En: *ISKO (5. 2001. Alcalá de Henares). La representación y organización del conocimiento: Metodologías, modelos y aplicaciones* [CD-ROM]. Alcalá de Henares: Universidad de Alcalá de Henares, 2001, p. 73-81.

LÓPEZ, Diego R. y Javier Massa. Dando forma al envase y, con ello, al contenido: Webber [documento HTML]. *Boletín de Red Iris*, n° 45, octubre 1998. Disponible en: <http://www.rediris.es/rediris/boletin/45/enfoque1.html>

LÓPEZ GUZMÁN, Clara. *Modelo para el desarrollo de bibliotecas digitales especializadas* [documento HTML]. México, D.F: UNAM, marzo de 2000. Disponible en: <http://www.bibliodgsca.unam.mx/tesis/tes7cllg/meta.html>

Bibliografía

LOZANO PALACIOS, Antonio. *Vocabulario inglés/español – español/inglés para los estudios de biblio-documentación*. 3ª ed. rev. y aum. Granada: Universidad de Granada, Facultad de Biblioteconomía y Documentación, octubre 1996. También disponible en Internet en: <http://eubd1.ugr.es/RIS/risweb.isa> (versión actualizada).

LOWE, David. Improving Web Search Relevance: Using Navigational Structures to Provide a Search Context [documento HTML]. En: *AusWeb2K: Australian World Wide Web Conference (6. 2000. Cairns)*. Disponible en: <http://ausweb.scu.edu.au/aw2k/papers/lowe/paper.html>

LUCAS, Marty. Demystifying Metadata [documento HTML]. *Mappa Mundi*, April 2000. Disponible en: <http://mappa.mundi.net/trip-m/metadata>

LUNATI, Gabriele. Bollicine di champagne...Strumenti per la ricerca ed il recupero dell'informazione su Internet. *Biblioteche oggi*, giugno 1997, XV, n° 5, p. 18-29. También disponible en Internet en: <http://www.burioni.it/forum/motori/motori.htm>

LUNATI, Gabriele. Strumenti per la ricerca ed il recupero dell'informazione su Internet. *Workshop Ellectronic Information in Libraries and Documentation Centres (5. 1996. Milán)*. En: Ricardo Ridi. *ESB Forum: Recensioni econtributi su cd-rom e altre fonti informative elettroniche*. Andrea Castagna e Patrizia Rivara (col.). Genova: E.S. Burioni Ricerche Bibliografiche, 1996. También disponible en Internet en: <http://www.burioni.it/forum/motori/par8.htm>

LUPOVICI, Catherine, Julien Masanès. *Metadata for Long-term Preservation* [documento PDF]. NEDLIB Project. La Hague: National Library of Netherlands, Koninklijke Bibliotheek, 29 de julio de 2000. Disponible en: <http://www.kb.nl/coop/nedlib/results/preservationmetadata.pdf>

LYNCH, Clifford. Searching the Internet¹. *Scientific American*, March 1997, p. 52-56.

LYNCH, Clifford. The Z39.50 Information Retrieval Standard. Part I, a Strategic View of its Past, Present and Future [documento HTML]. *D-Lib Magazine*, April 1997, vol. 3, n° 4. Disponible en: <http://www.dlib.org/dlib/april97/04lynch.html>

¹ La versión electrónica de este artículo estuvo disponible en la Web en el momento de realización de esta investigación en: <http://sciam.com/0397/issue/0397lynch.html> (consultado el 8 de julio de 1998). Actualmente la política de esta revista ha cambiado y para consultar la versión electrónica de números anteriores hay que pagar 5 US\$. Para acceder a este artículo por esa vía: <http://12.26.25.97/scripts/massets8010.dll/453272.NkMdCmHgGdBDfeApat> (consultado el 2 de mayo de 2001). No obstante, se puede acceder a la versión electrónica en francés de este paradigmático artículo a través de la siguiente referencia bibliográfica:

LYNCH, Clifford. La recherche d'informations [documento HTML]. *Pour la Science*, n° 235, mai 1997. También disponible en Internet en: <http://www.pourlascience.com/numeros/pls-235/internet/lynch.htm>

Bibliografía

LYNCH, Clifford and Héctor García-Molina. Interoperability, Scaling and Digital Libraries [documento HTML]. En: *Research Agenda. IITA Digital Libraries Workshop, August 22, 1995*. Stanford: University, 22 de agosto de 1995. Disponible en: <http://www-diglib.stanford.edu/diglib/pub/reports/itta-dlw/main.html> (consultado el 24 de julio de 1998).

M

MACDOUGALL, Susan. Signposts on the Information Superhighway: Indexes and Access. *Journal of Internet Cataloging*, 2000, vol. 2, n° 3/4, p. 61-79.

MAGÁN WALS, José Antonio. El concepto de biblioteca en la actualidad: bibliotecas reales frente a bibliotecas virtuales. En: *Tratado básico de biblioteconomía*. José Antonio Magán Wals, coord., 2ª ed. rev. y act. Madrid: Editorial Complutense, 1996, p. 19-46.

MAGUIRE, Carmel. Metadata in Australia. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, n° 1, p. 18-21.

MALAMUD, Carl, Marshall T. Rose. Maps, Space, and Other Metaphors for Meta-Data [documento HTML]. *Invisible Worlds, Inc, White Papers*, 3 de abril de 2000. Disponible en: <http://invisibleworlds.com/technology/metaphors.html> (consultado el 31 de octubre de 2000).

MALER, Eve, Jeanne El Andaloussi. *Developing SGML DTDs: From Text to Model To Markup*. New Jersey: Prentice Hall, 1996.

MANDEL, Carol. *Change and Continuity in Subject Authority Control* [documento HTML]. OCLC, rev. 22 de mayo de 1998. Disponible en: <http://www.oclc.org/oclc/man/9391ausy/mandle.htm>

MARC DTDs Document Type Definitions: Backgrounds and Development [documento HTML]. Washington, D.C.: Library of Congress, Network Development and MAR Standards Office, 22 de mayo de 1998. Disponible en: <http://www.loc.gov/marc/marcdtd/marcdtdback.html>

MARCHIORI, Massimo. The Limits of Web Metadata, and Beyond [documento HTML]. En: *International World Wide Web Conference (7. 1998. Brisbane)*, rev. 9 de septiembre de 1998. Disponible en: <http://www7.scu.edu.au/programme/fullpapers/1896/com1896.htm>

MARCOUX, Yves and Martin Sévigny. Why SGML? Why now? *Journal of the American Society for Information Science*, 1997, vol. 7, n° 48, p. 584-592.

Bibliografía

MARSHALL, Catherine C. Making Metadata: A Study of Metadata Creation for Mixed Physical-Digital Collection. En: *ACM Conference on Digital Libraries (3. 1998. Pittsburgh)*. Ian Witten, Rob Akscyn, and Frank M. Shipman, eds. New York: ACM, 1998, p. 162-171.

MARTIN, David. *The Information Broker Project* [documento HTML]. SRI International, 1 de noviembre de 1996. Disponible en: <http://www.ai.sri.com/~martin/papers/broker/techrep/memo/memo.html>

MARTÍN GALÁN, Bonifacio, David Rodríguez Mateos. Estructuración de la información mediante XML: un nuevo reto para la gestión documental. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao)*. *La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p. 113-123.

MARTÍN GALÁN, Bonifacio, Eva M^a Méndez Rodríguez. *Cómo citar y referenciar documentos electrónicos. Particularidades e iniciativas*. Trabajo presentado en el Doctorado en Documentación. Universidad Carlos III de Madrid, 1997, 53 p., (ejemplar original de autor).

MARTOS RUBIO, Ana. *Herramientas de búsqueda en Internet*. Madrid: Pearson Education, 2001.

[MASA MARÍN, Javier]. *Elementos del conjunto de metadatos [sic.] de Dublin Core: Descripción de referencia* [documento HTML]. RedIris, rev. 27 de noviembre de 2001. Disponible en: http://www.rediris.es/metadata/dublin_core_elements.es.html

MASA MARÍN, Javier. Gestión de recursos en la Red Académica. En: *ISKO (5. 2001. Alcalá de Henares)*. *La representación y organización del conocimiento: Metodologías, modelos y aplicaciones* [CD-ROM]. Alcalá de Henares: Universidad de Alcalá de Henares, 2001, p. 425-434.

[MASA MARÍN, Javier]. *Servicio de distribución de información indexada por la red (IRIS-INDEX)* [documento HTML]. RedIris, rev. 8 de julio de 1997. Disponible en: <http://www.rediris.es/si/iris-index/coordinacion/gt3/iris-index.html> (consultado el 24 de agosto de 1998).

MCCALLUM, Sally. Extending MARC for Bibliographic Control in the Web Environment: Challenges and Alternatives [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 1 de noviembre de 2000, 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/mccallum_paper.html

MCCRAY, Alexa T. And Marie E. Gallagher. Principles for Digital Library Development. *Communications of the ACM*, May 2001, vol. 44, n° 5, p. 49-54.

Bibliografía

MCKEMMISH, Sue, Adrian Cunningham and Dagmar Parer. *Metadata Mania* [documento HTML]. Melbourne: Monash University, School of Information Management & Systems, Records Continuum Research Group, rev. 5 de enero de 2000. Disponible en: <http://www.sims.monash.edu.au/rcrg/publications/recordkeepingmetadata/sm01.html> (consultado el 31 de agosto de 2000).

MEDEIROS, Norm. Making Room for MARC in a Dublin Core World. *Online*, 1999, vol. 23, nº 6, p. 57-60.

MEDEIROS, Norm. XML and the Resource Description Framework: The Great Web Hope [documento HTML]. *Online*, September 2000. Disponible en: <http://www.onlineinc.com/onlinemag/OL2000/medeiros9.html>

MEGGINSON, David. Re: Dissillusioned about Interoperability [mensaje de correo en lista de distribución]. *Xml-dev*, 8 de octubre de 1999; 08.11. Disponible en: <http://www.lists.ic.ac.uk/hypermail/xml-dev/xml-dev-Oct-1999/0101.html>

MELNIK, Sergey. *Storing RDF in a Relational Database* [documento HTML]. Stanford: Stanford University, Computer Science Department, 11 de mayo de 2000. Disponible en: <http://www-db.stanford.edu/~melnik/rdf/db.html>

MÉNDEZ RODRÍGUEZ, Eva M^a. *Metadatos y recuperación de información*. Tesina de doctorado, dirigida por Virginia Ortiz-Repiso Jiménez, presentada en septiembre de 1998, 210 p., (ejemplar original de autor).

MÉNDEZ RODRÍGUEZ, Eva M^a. Metadatos y tesauros: aplicación XML/RDF a los sistemas de organización del conocimiento en Intranets. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao). La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p. 211-219.

MÉNDEZ RODRÍGUEZ, Eva M^a. RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio. En: *Jornades Catalanes de Documentació (7. 1999. Barcelona)*. Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 1999, p. 487-498. También disponible en Internet en: <http://www.bib.uc3m.es/~mendez/publicaciones/7jc/rdf.htm>

MÉNDEZ RODRÍGUEZ, Eva M^a, José A. Merlo Vega. Localización, identificación y descripción de documentos web: tentativas hacia la normalización. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao). La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p. 221-231.

MÉNDEZ RODRÍGUEZ, Eva M^a, Murtha Baca. Acceso al patrimonio cultural en el mundo digital: estándares para la recuperación de información en museos virtuales. *Revista de Museología (Dossier: Museos del siglo XXI)*, nº 21, 1^{er} cuatrimestre 2001, p. 68-77.

Bibliografía

MERLO VEGA, José Antonio, Ángela Sorli Rojo. El uso de metainformación en los webs de las bibliotecas españolas. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao). La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p.154-164.

META Content Framework Using XML. Submitted to W3C 6 June 97 [documento HTML]. R.V. Guha, Tim Bray, eds. W3C, 24 de junio de 1997. Disponible en: <http://www.w3.org/TR/NOTE-MCF-XML-970624>

META Matters [documento HTML]. Canberra: National Library of Australia, 24 de mayo de 1999. Disponible en: <http://www.nla.gov.au/meta>

METABROWSER: a Web Browser for Metadata. Version 1.1 for Microsoft Windows [documento HTML]. Erindale: Metabrowser System, Spirit Consulting, 2001. Disponible en: <http://metabrowser.spirit.net.au/manual>

METADATA and Organizing Educational Resources on the Internet. Jane Greenberg, ed. New York, London, Oxford: The Harworth Information Press, 2000.

METADATA for Digital Preservation: the Cedars Project Outline Specification. Draft for Public Consultation [documento PDF]. Leeds: University of Leeds, Marzo 2000. Disponible en: <http://www.leeds.ac.uk/cedars/MD-STR~5.pdf>

METADATA Management System: Product Overview [documento PDF]. Kanata: Watchfire, September 2000. Disponible en: <http://www.watchfire.com/resources/meta-overview.pdf>

METADATA Manual [documento HTML]. Australian Virtual Engineering Library for Sustainable Development (AVELSD), rev. 5 de Julio de 2001. Disponible en: <http://avel.library.uq.edu.au/avelsd>

METADATA Resources [documento HTML]. Michael Day and Andy Powell, eds. Bath: University, UKOLN, 12 de noviembre de 2001. Disponible en: <http://www.ukoln.ac.uk/metadata>

METADATA Specification for BEP, Business Entry Point Management Branch of Department of Employment. Workplace Relations and Small Business, 1999. Disponible en: <http://about.business.gov.au/bep/agencies/provinfo/metadata/metadata.htm> (consultado el 28 de febrero de 2001).

METADATA Standards, Crosswalks, and Standard Organizations [documento HTML]. Charley Pennell. Queen Elizabeth II Library. Newfoundland: Memorial University of Newfoundland, rev. 14 de noviembre de 2001. Disponible en: <http://www.mun.ca/library/cat/standards.htm>

METADATA vs. Meta-data/Meta Data. [documento HTML]. En: *Metadata Company: Innovators in Data Management Technology*. Metadata Company, rev. 6 de marzo de 2000. Disponible en: <http://www.meta-data.com/word.htm>

Bibliografía

METADATA Workshop: Report of the Workshop Held in Luxembourg, 1 and 2 December, 1997 [documento PDF]. Makx Dekkers, ed. [s.l.]: European Commission, Directorate General XIII – E/4, February 1998.. Disponible en: <ftp://ftp.cordis.lu/pub/libraries/docs/metadata1.pdf>

METADATA Workshop: Report of the Workshop Held in Luxembourg, 26 June, 1998 [documento PDF]. Makx Dekkers, ed. [s.l.]: European Commission, Directorate General XIII – E/4, August 1998. Disponible en: <ftp://ftp.cordis.lu/pub/libraries/docs/metadata2.pdf>

METADATA Workshop: Report of the Third Workshop Luxembourg, 12 April 1999 [documento PDF]. Makx Dekkers, ed. [s.l.]: European Commission, Directorate General XIII, – E/2, June 1999. Disponible en: <ftp://ftp.cordis.lu/pub/libraries/docs/metadata3.pdf>

METAMANAGEMENT Concepts [documento HTML]. Metamanagement, rev. 26 de julio de 1999. Disponible en: <http://www.metamanage.com/doco/concepts.htm>

METATAGS: They're not Just for Search Engines: White paper [documento PDF]. Kanata: Watchfire, March 2000. Disponible en: <http://www.watchfire.com/resources/metatagswhite.pdf>

METAWEB Project: Analysis of Metadata Creation Tools [documento HTML]. Brisbane: University of Queensland, Distributed Systems Technologies Centre, rev. 31 de agosto de 1999. Diponible en: <http://www.dstc.edu.au/Research/Projects/metaweb/toolpost.html>

METTROP, Wouter. Intrenet Search Engines: Fluctuation in Document Accessibility. *Journal of Documentation*, September 2001, vol.57, n° 5, p. 623-651.

MILLÁN, José Antonio. *Internet y el español*. Madrid: Fundación Retevisión, 2001.

MILLER, Dick R. XML: Libraries' Strategic Opportunity [documento HTML]. *Library Journal NetConnect*, Summer 2000. Disponible en: <http://www.libraryjournal.com/xml.asp>

MILLER, Eric. An Introduction to the Resource Description Framework [documento HTML]. *D-Lib Magazine*, May 1998, vol. 4, n° 5. Disponible en: <http://www.dlib.org/dlib/may98/miller/05miller.html>

MILLER, Eric. Interoperability: What is it and Why Should I Want it? [documento HTML] *Ariadne*, 26-Jun-2000, issue 24. Disponible en: <http://www.ariadne.ac.uk/issue24/interoperability/intro.html>

MILLER, Eric. Making Progress: The Resource Description Framework (RDF). *Journal of Internet Cataloging*, 1999, vol. 1, n° 4, p. 53-58.

Bibliografía

MILLER, James S. W3C and Digital Libraries [documento HTML]. *D-Lib Magazine*, November 1996, vol. 2, n° 11. Disponible en: <http://www.dlib.org/dlib/november96/11miller.html>

MILLER, Paul. I Say What I Mean, But do I What I Say? [documento www] *Ariadne*, 22-March-2000, issue 23. Disponible en: <http://www.ariadne.ac.uk/issue23/metadata>

MILLER, Paul. Metadata for the Masses [documento HTML]. *Ariadne*, 11-September-1996, issue 5. Disponible en: <http://www.ariadne.ac.uk/issue5/metadata-masses>

MILSTEAD, Jessica and Susan Feldman. Metadata: Cataloging by any Other Name. *Online*, January/February 1999, vol. 23, n° 1, p. 24-31. También disponible en Internet en: <http://www.onlininc.com/onlinemag/OL1999/milstead1.html> (consultado el 3 de septiembre de 1999).

MILSTEAD, Jessica and Susan Feldman. Metadata: Projects & Standards. *Online*, January/February 1999, vol. 23, n° 1, p. 32-41. También disponible en Internet en: <http://www.onlininc.com/onlinemag/OL1999/milstead1.html> (consultado el 3 de septiembre de 1999).

MITCHELL, Steve, Margaret Mooney. INFOMINE: A Model Web-based Academic Virtual Library [documento HTML]. *Information Technology and Libraries*, March 1996, vol. 15, n° 1. Disponible en: <http://infomine.ucr.edu/pubs/italmine.html>

MODES of Cataloging Employed in the Cataloging Directorate [documento HTML]. Washington, DC.: Library of Congress, 27 de enero de 1997. Disponible en: <http://lcweb.loc.gov/catdir/catmodes.html>

MOEN, William E. Resource Discovery Using Z39.50: Promise and Reality [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 1 de noviembre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/moen_paper.html

MOEN, William E., Erin L. Stewart, and Charles R. McClure. The Role of Content Analysis in Evaluating Metadata for the U.S. Government Information Locator Service (GILS): Results from an exploratory Study [documento HTML]. En: *IEEE Meta-data Conference (2. 1997. Silver Spring)*. IEEE Computer Society, 1997, rev. 17 de febrero de 1997. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/wmoen/moen.htm> (consultado el 18 de junio de 1998).

MOODY, Glyn. A New Dawn [documento HTML]. *New Scientist*, 30 de mayo de 1998. Disponible en: <http://www.newscientist.com/ns/980530/xml.html> (consultado el 17 de julio de 1998).

Bibliografía

MOREIRO GONZÁLEZ, José Antonio, Eva M^a Méndez Rodríguez. Lenguaje natural e indexación automatizada. *Ciencias de la Información*, septiembre 1999, vol. 30, n^o 3, p. 11-24.

MORGAN, Eric Lease. Comparing Open Source Indexers [documento HTML]. *Infomotions Musings*, May 29, 2001. Disponible en: <http://www.infomotions.com/musings/opensource-indexers>

MORRIS, Sally. Metadata and Rights. *Vine (Theme issue: Metadata. Part 2)*, 1999, n^o 117, p. 30-48.

MOYA ANEGÓN, Félix de. *Los sistemas integrados de gestión bibliotecaria: Estructuras de datos y recuperación de información*. Madrid: ANABAD, 1995, 227 p.

MUDDAMALLE, Manikya Rao. Natural Language Versus Controlled Vocabulary in Information Retrieval: a Case Study in Soil Mechanics. *Journal of the American Society for Information Science*, August 1998, vol. 49, n^o 10, p. 881-887.

MUENNICH, Monika. Progress on the Multilingual Dictionary of Cataloging Terms and Concepts [documento HTML]. En: *IFLA Council and General Conference (67. 2001. Boston)* rev. 11 de noviembre de 2001. Disponible en: <http://www.ifla.org/IV/ifla67/papers/093-152ae.pdf>

MUIR, Adrienne. Legal Deposit of Digital Publications: A Review or Research and Development Activity. En: *Joint Conference on Digital Libraries (2001. Roanoke)*. New York: ACM, 2001, p. 165-173.

MURDOCH, Kaylene. Request for Information (RFI) for a Metadata Management Facility and Search Tool [correo electrónico en lista de distribución]. *www-rdf-interest*, 5 de agosto de 2001; 23:39. Disponible en: <http://lists.w3.org/Archives/Public/www-rdf-interest/2001Aug/0017.html>

MURPHY, Angela. Accessing the Visual Heritage: Metadata Construction at the Science & Society Picture Library. *Vine*, [1998], n. 107, p. 58-64.

MURPHY, Lisa D. Addressing the Metadata Gap: *ad hoc* Digital Documents in Organizations. En: *Text databases and document management: theory and Practice*. Amita Goyal Chin, ed. Hershey, etc.: Idea Group Publishing, 2001, p. 52-77.

MUXACH, Santi i Ana Lopo. Metadades a peu pla. *Item*, 1999, n^o 24, p. 99-134.

N

NAGEL, Lina. Using Standards in the Documentation of Museum Collections: Categories for the Description of Works of Art, Object ID, an Other Standardization Tools. *Spectra*, Spring 1999, vol. 26, issue 1, p. 36-39.

Bibliografía

NELSON, Michael L. *Smart Objects for Digital Libraries. Ph.D Dissertation*. Old Dominion University, August 2000 [documento RTF]. Disponible en: http://mln.larc.nasa.gov/~mln/phd/?method=display&pkg_name=dissertation.pkg&element_name=daily/phd.doc

NELSON, Michael L. and Kurt Maly. Buckets: Smart Objects for Digital Libraries. *Communications on the ACM*, May 2001, vol. 44, n° 5, p. 60-62.

NETGLOS - The Multilingual Glossary of Internet Terminology [documento HTML]. World Wide Languages Institute, 15 de marzo de 1997, rev. 21 de febrero de 1998. En: <http://wwli.com/translation/netglos/glossary/glossary.html>

NICHOLS, Stephen G., Abby Smith. *The Evidence in Hand: Report of the Task Force on the Artifact in Library Collections*. Washington: CLIR, November 2001. También disponible en Internet en: <http://www.clir.org/pubs/reports/pub103/pub103.pdf>

NICHOLSON, Dennis, et al. *Cataloguing the Internet: CATRIONA Feasibility Study. Report To The British Library Research & Development Department* [documento HTML]. The British Library Board, BUBL, 1995. Disponible en: <http://bubl.ac.uk/org/catriona/cat1rep.htm>

NICHOLSON, Dennis, Martin Smith. Electronic Resource Creation and Management at Scottish Universities: Survey Results and Demonstrator Service Progress [documento HTML]. *Ariadne*, 26-June-1998, issue 14. Disponible en: <http://www.ariadne.ac.uk/issue14/catriona>

NICHOLSON, Scott. Indexing and Abstracting on the World Wide Web: An Examination of Six Web Databases. *Information Technology and Libraries*, June 1997, vol. 16, n° 2, p. 73-81.

NIELSEN, Jakob. *Usabilidad. Diseño de sitios Web*. Madrid: Pearson Educación, 2000.

NISO Draft Standard. Data Dictionary: Technical Metadata for Digital Still Images [documento PDF]. NISO, 5 de julio de 2000. Disponible en: <http://www.niso.org/pdfs/DataDict.pdf>

[NOERR, Peter]. *The Digital Library Toolkit.2nd* ed. Palo Alto: Sun Microsystems, March 2000.

Les NOUVELLES technologies dans les bibliothèques. Michèle Rouhet, dir. Paris: Électre – Éditions du Cercle de la Librairie, 1996.

NÚÑEZ, Lluïsa. *Metadades: informe per al CBUC* [documento PDF]. Barcelona: Biblioteca Digital de Catalunya, 2000. Disponible en: <http://www.cbuc.es/5digital/meta0007.pdf>

O

O'NEILL, Edward. Characteristics of Web Accessible Information. *IFLA Journal*, 1998, vol. 24, nº 2, p. 114-116. También disponible en Internet en: <http://www.nlc-bnc.ca/ifla/IV/ifla63/63onee.htm> (consultado el 8 de julio de 1998).

OCLC Offers OpenURL Option for Libraries [documento HTML]. Dublin: OCLC, 25 de octubre de 2001. Disponible en: <http://www.oclc.org/oclc/press/20011025.shtm>

ODDY, Pat. *Future Libraries: Future Catalogues*. London: Library Association Publishing, 1996, 179 p.

ODER, Norman. Cataloging the Net: Can We Do It? *Library Journal*, October 1998, vol. 123, issue 16, p. 47-51.

OGBUJI, Uche. An Introduction to RDF: Exploring the Standard for Web-Based Metadata [documento HTML]. *IBM Web Architecture articles*, December 2000. Disponible en: <http://www-106.ibm.com/developerworks/webservices/library/w-rdf>

OLVERA LOBO, M^a Dolores. Rendimiento de los sistemas de recuperación de información en la Web: evaluación de servicios de búsqueda (*search engines*). *Revista Española de Documentación Científica*, 2000, vol. 23, nº 3, p. 302-316.

ONTALBA Y RUIPÉREZ, José Antonio. Bibliotecas digitales españolas: análisis de las primeras realidades. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao). La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p. 399-409.

ON the Size of the World Wide Web [documento HTML]. *Pandia*, 14 de octubre de 2001. Disponible en: <http://www.pandia.com/sw-2001/57-websize.html>

The OPEN Archives Initiative Protocol for Metadata Harvesting. Protocol Version 1.1 of 2001-07-02 [documento HTML]. Herbert Van de Sompel, Carl Lagonze, eds. OAI, 20 de junio de 2001. Disponible en: http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html

OPPENHEIM, Charles, et al. The Evaluation of WWW Search Engines. *Journal of Documentation*, March 2000, vol. 56, nº 2, p. 190-211.

OPPENHEIM, Charles and Daniel Smithson. What is the Hybrid Library? *Journal of Information Science*, 1999, vol. 25, nº 2, 97-112.

ORTEGA Y GASSET, José. *La misión del bibliotecario*. Málaga: Asociación Andaluza de Bibliotecarios, 1994.

P

PACK, Thomas. Go Forth and Syndicate: Five Questions to Ask When Getting Started. *Econtent*, August 2001, vol. 24, nº 6, p. 24-30.

PANCHYSHYN, Roman S., France Bouthillier. Cataloguer le cyberspace: le défi des ressources électroniques. *Documentation et Bibliothèques*, juillet-septembre 1997, vol. 43, nº 3, p. 137-147.

PAPAGEORGIU, H., Maria Vardaki, Fragkiskos Pentaris. Recent Advances on Metadata. *Computational Statistics*, 2000, vol. 15, nº 1, p. 89-97.

PATEL, Manjula. *SCHEMAS: Second Workshop Report: Publishing and sharing your metadata application profile* [documento PDF]. Schemas Forum, 26 de enero de 2001. Disponible en: <http://www.schemas-forum.org/workshops/ws2/wks2.pdf>

PEDRAZA GARCÍA, Manuel José. Aproximación a la catalogación de documentos electrónicos de acceso remoto. *Scire*, enero-junio 1997, vol. 3, nº 1, p. 99-125.

PEDRAZA GARCÍA, Manuel José. Las ISBD (ER): Adecuación normativa a un panorama cambiante. *Scire*, enero-junio 1998, vol. 4, nº 1, p. 45-55.

PEIS, Eduardo, Felix de Moya. Sgml y servicios de información. *El profesional de la información*, junio 2000, vol. 9, nº 6, p. 4-17.

PEIS, EDUARDO, Juan Carlos Fernández Molina. Modelos metadatos para la gestión de documentos electrónicos. En: *ISKO (5. 2001. Alcalá de Henares). La representación y organización del conocimiento: Metodologías, modelos y aplicaciones* [CD-ROM]. Alcalá de Henares: Universidad de Alcalá de Henares, 2001, p.148-157.

PEIXOTO BAX, Marcello. Introdução às linguagens de marcas [documento HTML]. Belo Horizonte: Universidade Federal de Minas Gerais, 14 de abril de 2000. Disponible en: http://www.eb.ufmg.br/Bax/Disciplinas/BiblioDigi/Introdução%20às%20Markup%20Languages_submetido.htm

PEPPER, Steve. *The TAO of Topic Maps: Finding The Way in the Age of Infoglut* [documento PDF]. Ontopia, 2001. Disponible en: <http://www.ontopia.net/topicmaps/materials/tao.pdf> También disponible en: <http://www.gca.org/papers/xmleurope2000/pdf/s11-01.pdf>

PHILLIPS, Hugh. *Metadata Tools for Geospatial Data* [documento HTML]. Wisconsin Land Information Clearinghouse, rev. 11 de noviembre de 2001. <http://badger.state.wi.us/agencies/wlib/sco/metatool/mtools.htm>

PHILLIPS, Lee Anne. *Descubre HTML 4*. Madrid, etc.: Prentice Hall, 1998.

Bibliografía

PHILLIPS, Lee Anne. *Using XML: Special Edition*. Bestseller ed. Indianapolis: Que, 2000.

PISCITELLI, Alejandro. *Ciberculturas. En la era de las máquinas inteligentes*. Buenos Aires: Paidós, 1995.

PITTS, Natanya. *XML*. Madrid: Anaya Multimedia, 1999.

PITTS-MOULTIS, Natanya, Cheryl Kirk. *XML Black Book*. Albany, etc.: Coriolis, 1999.

PODIO, Fernando L., Willian Vollrath and Ben Kobler. Media Error Monitoring and Reporting Information (MEMRI). Metadata for Intelligent Digital Data Storage Devices [documento HTML]. En: *IEEE Meta-data Conference (2. 1997. Silver Spring)*. IEEE Computer Society, 1997, rev. 17 de febrero de 1998. Disponible en: <http://computer.org/conferen/proceed/meta97/papers/fpodio/fpodio.htm> (consultado el 18 de junio de 1998).

POLLOCK, Annabel, Andrew Hockley. What's Wrong with Internet Searching [documento HTML]. *D-Lib Magazine*, March 1997, vol. 3, nº 3. Disponible en: <http://www.dlib.org/dlib/march97/bt/03pollock.html>

POLYDORATOU, Panayiota, David Nicholas. Familiarity with and Use of Metadata Formats and Metadata Registries amongst Those Working in Diverse Professional Communities Within the Information Sector. *ASLIB Proceedings*, 2001, vol. 53, nº 8, p. 309-324.

POUCHARD, Line. Cataloging for Digital Libraries: The TEI Scheme and the TEI Header [documento HTML]. *Katharine Sharp Review*, Winter 1998, issue 6. Disponible en: <http://mirrored.ukoln.ac.uk/lis-journals/review/review/6>

POWELL, Andy, Ann Apps. Encoding OpenURLs in Dublin Core Metadata [documento HTML]. *Ariadne*, 23-March-2001, issue 27. Disponible en: <http://www.ariadne.ac.uk/issue27/metadata>

POWELL, Andy, Lorcan Dempsey. Dublin Core and Metadata: a Tutorial. En: *Metadata Workshop, Luxembourg, 15 December 1997* [documento HTML]. Bath: University, UKOLN, rev. 28 de febrero de 1998. Disponible en: <http://hosted.ukoln.ac.uk/ec/metadata-1997/tutorial/presentation>

POWELL, Andy. *Dublin Core in XML: a Guide for Implementors*. 3rd draft. [documento HTML]. Bath: University, UKOLN Metadata Group, rev. 17 de octubre de 2001. Disponible en: <http://www.ukoln.ac.uk/metadata/resources/dc/dc-xml-guidelines>

POWELL, Andy. *Metadata UKOLN Software Tools* [documento HTML]. Bath: University, UKOLN, rev. 9 de julio de 2001. Disponible en: <http://www.ukoln.ac.uk/metadata/software-tools>

Bibliografía

PRICE, Roger. Beyond SGML. En: *ACM Conference on Digital Libraries (3. 1998. Pittsburg)*. Ian Witten, Rob Akscyn, Frank M. Shipman, eds. New York: ACM, 1998, p. 172-181.

PRICEWATERHOUSECOOPERS. *Metadata Watch Report #1 Top-Level Synthesis* [documento PDF]. SCHEMAS, 8 de mayo de 2000. Disponible en: <http://www.schemas-forum.org/metadata-watch/1.pdf>

PROTHMAN, Bill. Meta Data: Managing Needles in the Proverbial Haystacks [documento PDF]. *IEEE Potentials*, february-march 2000. Disponible en: <http://www.ieee.org/membership/students/potentials/febmar2000/pdf/pages20.pdf>

Q

QUIN, Jian. Representation and Organization of Information in the Web Space. *Informing Science*, 2000, vol. 3, nº 2, p. 83-88. También disponible en Internet en: <http://inform.nu/Articles/Vol3/v3n2p83-88.pdf>

R

The RDF Thesaurus Description Standard [documento HTML]. California: Environmental Resources Evaluation System, California Resources Agency, rev. 26 de septiembre de 2000. Disponible en: <http://ceres.ca.gov/thesaurus/RDF.html>

REPORT of a Formal Meeting Held at Rauol Wallenberg Centre Lund, Sweden. 4 April 2001 [documento PDF]. Cordis, Digicult, abril de 2001. Disponible en: <ftp://ftp.cordis.lu/pub/ist/docs/digicult/lund-report.pdf>

REQUEST For Information (RFI): Metadata Management Facility & Search Tool [documento HTML]. New Zeland e-government, State Services Commission, 2001. Disponible en: <http://e-government.govt.nz/vendor/rfi-list.html> (consultado el 6 de agosto de 2001).

RESEARCH and Advanced Technology for Digital Libraries. European Conference ECDL (4. 2000. Lisbon). José Borbinha; Thomas Baker, eds. Berlin: Heidelberg; New York, etc: Springer, 2000.

RESEARCH in Metadata for Computers Model [documento HTML] Scott Crosier, dir. Santa Barbara: University of California. Geography Department, rev. 14 de junio de 2001. Disponible en: <http://www.geog.ucsb.edu/~scott/metadata>

RICHARD, Trina. *Managing Internet Resources Cataloguing* [documento HTML]. Toronto: Ryerson University, 15 de junio de 1998, 3ª rev., marzo de 1999. Disponible en: <http://www.ryerson.ca/~trichard/mir.html>

Bibliografía

- RIEMER, John. Adding 856 Fields to Authority Records: Rationale and Implications. *Cataloging & Classification Quarterly*, 1998, vol. 26, nº 2, p. 5-9.
- RIJSBERGEN, C. J. van. *Information Retrieval*. 2nd ed. London: Butterworths, 1979 [documento HTML]. Glasgow: University of Glasgow, Computing Science, rev. 18 de septiembre de 1996. Disponible en: <http://www.dcs.glasgow.ac.uk/Keith/Preface.html>
- RIVAS, Yolanda. "Surfea" en el cyberspanglish [documento HTML]. *El Mundo*, viernes 19 de junio de 1998. Disponible en: <http://www.el-mundo.es/navegante/diario/98/junio/19/cyberspanglish.html>
- ROBBINS, Fran. An Exploration of the Application of Classification Systems as a Meted of Resource Delivery on the World Wide Web. *Cataloguing Australia*, 1999, vol. 25, nº 1/4, p. 60-65.
- RODRÍGUEZ I GAIRÍN, Josep Manuel. Valoración del impacto de la información en Internet: Altavista, el «Citation Index» de la Red. *Revista Española de Documentación científica*, 1997, vol. 20, nº 2, p. 175-181
- RODRÍGUEZ MUÑOZ, José Vicente. Documentos electrónicos y normalización: información y conocimiento. Perspectivas de futuro. *Scire*, enero-junio 1997, vol. 3, nº 1, p. 139-151.
- ROLE, François. *Panorama des travaux en cours dans le domaine des métadonnées: Rapport de recherche n° 3628*. [documento PDF]. Paris: INRIA, février 1999. Disponible en: <ftp://ftp.inria.fr/INRIA/publication/publi-pdf/RR/RR-3628.pdf>
- ROMANO REYNOLDS, Regina. Partnership to mine unexploited sources of metadata [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 17 de octubre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/reynolds_paper.html
- ROSA, Antonio de la. Instrumentos terminológicos en el WWW: XML. *El profesional de la Información*, octubre 1999, vol. 8, nº 10, p. 14-36.
- ROSA, Antonio de la. Xml orientado a objetos. *El profesional de la información*, septiembre 1999, vol. 8, nº 9, p. 4-23.
- ROSZKOWSKI, Michael and Christopher Lukas. A Distributed Architecture for Resource Discovery Using Metadata [documento HTML]. *D-Lib Magazine*, June 1998, vol. 4, nº 6. Disponible en: <http://www.dlib.org/dlib/june98/scout/06roszkowski.html>
- ROUMIEUX, Olivier. Balise Meta: indexer ses pages Web. *Archimag*, juillet/août 1998, nº 116, p. 41.

Bibliografía

ROUMIEUX, Olivier. La bibliothèque virtuelle. *Archimag*, Février 1998, nº 111, p. 23-32.

ROUMIEUX, Olivier. François Chahuneau: «tout le monde a intérêt au développement du XML». *Archimag*, abril 1998, nº 113, p. 16.

ROVIRA SOLER, José Carlos. *Sobre la Biblioteca Virtual "Miguel de Cervantes"* [documento HTML]. Alicante: Biblioteca Virtual Miguel de Cervantes, 27 de julio de 2001. Disponible en: <http://www.cervantesvirtual.com/aniversario2/balance.shtml> (consultado el 30 de junio de 2001).

ROVIRA SOLER, José Carlos. Sobre la experiencia en nuevas tecnologías de la información de la universidad de Alicante y la Biblioteca Virtual Miguel de Cervantes [documento RTF]. En: *Congreso Internacional de la Lengua Española (2. 2001. Valladolid)*. Disponible en: http://cvc.cervantes.es/obref/congresos/valladolid/fronteras/ponencia/rovira_j.doc (consultado el 17 de octubre de 2001). También disponible en: http://cervantesvirtual.com/noticias/congresolengua/ponencia_rovira.htm

ROWLATT, Mary, et al. *SEAMLESS: an Organisational and Technical Model for Seamless Access to Distributed Citizen's Information*. London: Library and Information Commission, 2000.

RUSBRIDGE, Chris. Towards the Hybrid Library [documento HTML]. *D-Lib Magazine*, July/August 1998, vol. 4, nº 7/8. Disponible en: <http://www.dlib.org/dlib/july98/rusbridge/07rusbridge.html>

RUSCH-FEJA, Diann. Subject-Oriented Collection of Information Resources from the Internet: A Clearinghouse Concept to Support Scientist in a German Research Institute. *Libri*, 1997, vol. 47, p. 1-24.

RUSSELL, Rosemary and Michael Day. *HILT: High-Level Thesaurus project: Automated and Manual Approaches to the Provision of Thesauri and Subject Vocabularies* [documento HTML]. Bath: University, UKOLN, 5 de octubre de 2001. Disponible en: <http://www.ukoln.ac.uk/metadata/hilt/interfaces>

RUST, Godfrey. Metadata: The Right Approach. An Integrated Model for Descriptive and Rights Metadata in E-commerce [documento HTML]. *D-Lib Magazine*, July/August 1998, vol. 4, nº 7/8. Disponible en: <http://www.dlib.org/dlib/july98/rust/07rust.html>

RUST, Godfrey, Mark Bride. *The <Indecs> Metadata Framework: Principles, Model and Data Dictionary* [documento PDF]. <Indecs>TM Framework, Ltd., 2 de junio de 2000. Disponible en: <http://www.indecs.org/pdf/framework.pdf>

RUSTY HAROLD, Elliotte. *XMLTM Bible*. New York, etc.: IDG Books Worldwide, 1999.

S

SAÀDANI, Lalthoum, Suzanne Bertrand-Gastaldy. La représentation dans Internet des connaissances d'un domaine. *Documentation et bibliothèques*, janvier-mars 2000, vol. 46, n° 1, p. 27-42.

SAHAMI, Mehran, Salim Yusufali, Michelle Q. W. Baldonado. SONIA: A Service for Organizing Networked Information Autonomously. En: *ACM Digital Libraries (3. 1998. Pittsburg)*. Ian Witten, Rob Akscyn, Frank M. Shipman, III, eds. New York: ACM, 1998, p. 200-208.

SAMPEDRO LOSADA, Pedro José. *Anglicismos, barbarismos, neologismos y "falsos amigos" en el lenguaje informático* [documento HTML]. ATI, Grupo de Lengua e Informática, rev. 6 de junio de 2000. Disponible en: <http://www.ati.es/gt/lengua-informatica/externos/sampedr1.html>

SAN SEGUNDO MANUEL, Rosa. Organización del conocimiento en Internet: metadatos bibliotecarios Dublin Core. En: *Jornadas Españolas de Documentación (6. 1998. Valencia)*. Valencia: Fesabid, 1998, p. 805-818.

SÁNCHEZ ARCE, M^a Vanessa, Tomás Saorín Pérez. Las comunidades virtuales y los portales como escenarios de gestión documental y difusión de información. *Anales de Documentación*, 2001, n° 4, p. 215-228.

SANTAMARÍA, Fernando. XML (eXtensible Markup Language): nuevo estándar para la descripción de documentos en la World Wide Web. En: *Jornadas Españolas de Documentación (6. 1998. Valencia)*. Valencia: Fesabid, 1998, p. 819-826.

SAORÍN PÉREZ, Tomás. El concepto de biblioteca híbrida. *Tejuelo*, 2001, n° 1 [documento PDF]. Murcia: Universidad, Grupo de Tecnologías de la Información, rev. 1 de octubre de 2001. Disponible en: <http://gti1.edu.um.es:8080/saorin/articulos/BHibrida-Tejuelo2001.PDF>

SAORÍN PÉREZ, Tomás *Los Portales Bibliotecarios: Gestión automatizada de bibliotecas y Sistema de Información Público en Internet* [documento PDF]. Murcia: Universidad, Grupo de Tecnologías de la información. Disponible en: <http://gti1.edu.um.es:8080/portales/FGSR-portales-Bibliotecas.PDF>

SCHATZ, Bruce, et al. Federated Search of Scientific Literature. *Computer*, February 1999, p. 51-59.

SCHESCHY, Virginia M. Cataloging Procedures on the Web: the Greatest Thing since MARC. *Cataloging & Classification Quarterly*, 1998, vol. 26, n° 2, p. 11-23.

SCHREIBMAN, Susan. Metainformation Strategies for Electronic Resources. En: *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (12. 2000. Glasgow)*.

Bibliografía

ALLC/ACH 2000 Conference Abstracts. Glasgow: University of Glasgow, 2000, p. 114-115.

SCHWARTZ, Candy. *Definitions-LIS 530Z: Digital Libraries* [documento XHTML]. Boston: Simmons College, rev. 30 de marzo de 2000. Disponible en: <http://web.simmons.edu/~schwartz/530-defs.html> (consultado el 10 de febrero de 2001).

SCHWARTZ, Candy. *Metadata Portals and Multistandard Projects* [documento XHTML]. Boston: Simmons College, rev. 31 de agosto de 2001. Disponible en: <http://web.simmons.edu/~schwartz/meta.html>

SCHWARTZ, Ray. Metadata and the Web: Uniform Resource Identifiers and the Effort to Bring “Bibliographic Control” to the Web: An Overview of Current Progress. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, nº 1, p. 12-14.

The SCOUT Report selection criteria [documento HTML]. En: *The Scout Report Surf Smarter*. [Wisconsin]: University of Wisconsin-Madison, Internet Scout Report, 1998, rev. 22 de junio de 1998. Disponible en: <http://scout.cs.wisc.edu/scout/report/criteria.html>

SEINER, Robert S. *Questions Metadata Can Answer* [documento PDF]. Computer Associates Products, CAI, 14 de julio de 2000. Disponible en: http://www.cai.com/products/decisionbase/questions_metadata_can_answer.pdf

SENSO, José A., Antonio de la Rosa. Especificaciones XML aplicadas a la Documentación. En: *Anuario de Biblioteconomía, Documentación e Información. Bibliodoc 1999*. Barcelona: Col.legi Oficial de Bibliotecaris-Documentalistes de Catalunya, 1999, p. 133-152.

SERENDIPITY homepage [documento HTML]. Ohio: SouthEastern Ohio Regional Freenet Ohio University-Governor’s Scholars, 18 de julio de 1996, rev. 26 de julio de 1996. Disponible en: <http://www.seorf.ohiou.edu/~xx045/hide/serendip/serendip.htm> (consultado el 18 de julio de 1998).

SHAFER, Keith. *Mantis, A Flexible Cataloging Toolkit* [documento HTML]. OCLC, 22 de octubre de 1998. Disponible en: <http://orc.rsch.oclc.org:6464/toolkit.html>

SHAFER, Keith. Scorpion Helps Catalog the Web. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, nº 1, p. 28-29.

SHARP, Kate. Biblioteconomía en Internet: papeles tradicionales en un entorno nuevo. *IFLA Council and General Conference (66. 2000. Jerusalem)* [documento HTML]. IFLA, 1 de agosto de 2000. Disponible en: <http://www.ifla.org/IV/ifla66/papers/005-120s.htm>

Bibliografía

SHIRI, Ali Asghar and Crawford Revie. Thesauri on the Web: Current Developments and Trends. *Online Information Review*, 2000, vol. 24, nº 2, p. 273-279.

SMITH, Terence R. Meta-Information Environment of Digital Libraries [documento HTML]. *D-Lib Magazine*, July-August 1996, vol. 2, nº 7/8. Disponible en: <http://www.dlib.org/dlib/july96/new/07smith.html> También disponible en: <http://www.alexandria.ucsb.edu/public-documents/dlib>

La SOCIEDAD de la Información: política, tecnología e industria de los contenidos. Mercedes Caridad Sebastián, coord. Madrid: Fundación Ramón Areces, 1999.

SOCIETY OF AMERICAN ARCHIVIST. *EAD. Descripción Archivística Codificada: directrices de aplicación. Versión 1.0*. Fundación Histórica Tavera, ed., trad. [Alicante]: Fundación Histórica Tavera, 2000.

SOERGEL, Dagobert. Enriched Thesauri as Networked Knowledge Bases for People and Machines [documento PDF]. En: *CENDI Conference Controlled Vocabulary and The Internet (1999. Bethesda)*. CENDI, Subject Analysis and Retrieval Working Group, 1999. Disponible en: <http://www.dtic.mil/cendi/presentations/cendisoergel.pdf>

SOLBRIG, Harold R. Metadata and the Reintegration of Clinical Information: ISO 11179. *MD Computing*, May/June 2000, vol. 17, nº 3, p. 25-28.

SOLUCIONES alternativas al empleo de robots: Indización distribuida [documento HTML]. Granada: Universidad, Facultad de Biblioteconomía y Documentación, 1996, rev. 17 de noviembre de 1997. Disponible en: <http://eubd1.ugr.es/temp/serbydor/g9/know4.htm>

SORLI ROJO, Ángela, José Antonio Merlo Vega. Bibliotecas Digitales (I): Colecciones de libros de acceso público. *Revista Española de Documentación Científica*, 2000, vol. 23, nº 1, p. 91-103.

SOWA, John F. *Ontology, Metadata and Semiotics* [documento HTML]. Best Web Internet, rev. 12 de abril de 2001. Disponible en: <http://www.bestweb.net/~sowa/peirce/ontometa.htm>

SPAM [documento HTML]. En: *Clear Presence: Web Promotion Workshop*. San Luis Obispo, CA: The Grid, rev. 9 de enero de 1998. Disponible en: <http://www.thegrid.net/clear/spam.htm> (consultado el 6 de agosto de 1998).

SPAMMING [documento HTML]. En: *Computer Ethics*. North Carolina State University, rev. 28 de mayo de 1998. Disponible en: http://www2.ncsu.edu/eos/info/computer_ethics/www/spamming/index.html (consultado el 6 de agosto de 1998).

SPERBERG-MCQUEEN, C. M. *XML and What It Will Mean for Libraries* [documento HTML]. Chicago: University of Illinois, 30 de junio de 1998. Disponible en: <http://www.uic.edu/~cmsmcq/talks/teidlf1.html>

Bibliografía

SPERBERG-MCQUEEN, C. M., Claus Huitfeldt, Allen Renear. Meaning and Interpretation of Markup. En: *Joint International Conference of the Association for Literary and Linguistic Computing and the Association for Computers and the Humanities (12. 2000. Glasgow)*. ALLC/ACH 2000 Conference Abstracts. Glasgow: University of Glasgow, 2000, p. 27-30.

SREENIVASULU, V. The Role of a Digital Librarian in the Management of Digital Information System (DIS). *The Electronic Library*, 2000, vol. 18, nº 1, p. 12-20.

ST. LAURENT, Simon. Describing Your Data: DTDs and XML Schemas [documento HTML]. *Xml.com*, 1 de diciembre de 1999. Disponible en: <http://www.xml.com/pub/1999/12/dtd/index.html>

ST. LAURENT, Simon, Robert Biggar. *Inside XML DTDs*. New York, etc.: McGraw Hill, 1999.

ST. PIERRE, Margaret. Organisational Profiles: Blue Angel Introduces MetaStar [documento HTML]. *Exploit interactive*, July 1999. Disponible en: <http://www.exploitlib.org/issue2/blueangeltech>

ST. PIERRE, Margaret. Z39.50 and Semantic Interoperability [documento HTML]. En: *News and Info*. Blue Angel Technologies, 8 de enero de 1998. Disponible en: <http://www.blueangel.com/NewsAndInfo/paperz39.htm> (consultado el 18 de junio de 1998).

ST. PIERRE, Margaret, James Restivo. The MetaStar Suite of Metadata Tools [documento HTML]. En: *IEEE Metadata Conference (2. 1997. Silver Spring)*. IEEE Computer Society, 1997. Disponible en: <http://www.computer.org/proceedings/meta97/papers/mstpierre/mstpierre.html>

ST. PIERRE, Margaret, James Restivo. An Integrated Tool Suite for Managing Metadata. En: *Proceedings of the Earth observation & Geo-Spatial Web and Internet workshop*. *Salzburger Geographische Materialien*, 1998, vol. 27. También disponible en Internet en: <http://www.sbg.ac.at/geo/eogeo/authors/stpierre/stpierre.htm>

ST. PIERRE, Margaret, William P. LaPlant, Jr. Issues in Crosswalking, Content Metadata Standards [documento HTML]. *NISO Standards*. Bethesda: NISO, 15 de octubre de 1998. Disponible en: <http://www.niso.org/crswalk.html> (consultado el 5 de octubre de 2000).

STAAB, S., et al. Semantic Community Web Portals [documento HTML]. En: *International World Wide Web Conference (9. 2000. Amsterdam): The Web, next generation*, rev. 20 de noviembre de 2000. Disponible en: <http://www.www9.org/w9cdrom/134/134.html>

SUE, Hoylen. *XER - A Bridge between ASN.1 and XML* [documento HTML]. Brisbane: University of Queensland, Distributed Systems Technologies Centre, 12 de

Bibliografía

enero de 1998, rev. 23 de junio de 1999. Disponible en: <http://www.dstc.edu.au/Research/Projects/xer/xer-2.html>

SUMMERS, Ed. Chaos—Show Me the Metadata!. *Against the Grain*, February 1999, p. 79-82.

SUN MICROSYSTEMS. *Augmented Metadata in XHTML: Sun Microsystems Working Draft 21 June 2001* [documento HTML]. Murray Altheim, Sean B. Palmer, eds. Sun Microsystems, 21 de junio de 2001. Disponible en: <http://www.doctypes.org/meta/NOTE-xhtml-augmeta.html>

SUTTON, Stuart A. Conceptual Design and Deployment of Metadata Framework for Educational Resources on the Internet. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, n° 13, p.1182-1192.

SUTTON, Stuart A. GEM: Using Metadata to Enhance Internet Retrieval by K-12 Teachers. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, n° 1, p. 21-24.

SWICK, Ralph R. *Errata in REC-rdf-syntax-19990222* [documento HTML]. W3C, 31 de mayo de 2001. Disponible en: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/errata>

T

TALAVERA IBARRA, Ana M^a. De la biblioteca virtual a la biblioteca real: consideraciones en torno a la organización de la información en la era electrónica. En: *Abla 2000* [CD-ROM], 13 p.

TALMACS, Kerrie. Scott and the Race for the South Pole: the Triumph of Metadata. En: *National Cataloguing Conference (13. 1999. Brisbane): Charting the Information Universe* [documento HTML]. Sydney: University of New South Wales Library, 14 de octubre de 1999, rev. 1 de noviembre de 1999. Disponible en: <http://www.library.unsw.edu.au/~eirg/conf/natcat99.html>

TAYLOR, Arlene G., Patrice Clemson. Access to Networked Documents: Catalogs? Search Engines? Both? [documento HTML]. En: *OCLC Internet Cataloging Colloquium (1996. San Antonio)*. OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/taylor.htm>

TAYLOR, Chris. *An Introduction to Metadata* [documento HTML]. Brisbane: University of Queensland, rev. 1 de abril de 1999. Disponible en: <http://www.library.uq.edu.au/iad/ctmeta4.html>

TEASDALE, Guy. *L'édition savante à l'ère de la bibliothèque virtuelle: publication d'un livre en SGML sur le World Wide Web*. EBSI, Université de Montréal, 1996.

Bibliografía

[documento HTML]. Québec: Université de Laval, 1998. Disponible en: http://www.bibl.ulaval.ca/info/pagepers/teasdale/td_guy.htm

TEASDALE, Guy. De la nécessité des métadonnées sur Internet et les Intranets [documento HTML]. *La Lettre du bibliothécaire québécois*. Janvier-mars 2000, n° 21. Disponible en: <http://www.sciencepresse.qc.ca/lbq/lbq21.6.html>

TEI Guidelines for Electronic Text Encoding and Interchange (P3) [documento HTML]. Virginia: University of Virginia, Electronic Text Center, [1998?]. Disponible en: <http://etext.virginia.edu/TEI> (consultado el 24 de agosto de 1998).

TEI U5: Encoding for Interchange: an Introduction to the TEI [documento HTML]. TEI Consortium, 1995, rev. 21 de enero de 2001. Disponible en: <http://www.hcu.ox.ac.uk/TEI/Lite>

TENNANT, Roy. XML: The Digital Library Hammer [documento HTML]. *Library Journal, Infotech: Digital libraries*, 15 de marzo de 2001. Disponible en: http://www.libraryjournal.com/articles/infotech/digitallibraries/20010315_17960.asp

TERCEIRO, José B. y Gustavo Matías. *Digitalismo: el nuevo horizonte sociocultural*. Madrid: Taurus, 2001.

THIELE, Harold. The Dublin Core and Warwick Framework: a Review of the Literature, March 1995 – September 1997 [documento HTML]. *D-Lib Magazine*, January 1998, vol. 4, n° 1. Disponible en: <http://www.dlib.org/dlib/january98/01thiele.html>

THOMAS, Charles F. and Linda S. Griffin. Who Will Create the Metadata for Internet? [documento HTML]. *First Monday*, December 1998, vol. 3, n° 12. Disponible en: http://www.firstmonday.dk/issues/issue3_12/thomas/index.html

THOMAS, Sarah E. The Catalog as Portal to the Internet [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000. Washington)*. Sao Paulo: Universidade de Campinas, 16 de noviembre de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/thomas_paper.html

THOMAS-OGBUJI, Chimezie. The Future of Natural-Language Processing: Seeking a Practical Solution to the Limitations of NLP Methodologies and Software [documento HTML]. *Unix insider*, rev. 3 de enero de 2001. Disponible en: http://unixinsider.com/unixinsideronline/swol-12-2000/swol-1229-ontology_p.html (consultado el 10 de enero de 2001).

THORNELLY, Jennie. The How of Metadata: Metadata Creation and Standards. 1999, *Cataloguing Australia*, vol. 25, n° 1/4, p. 25-35.

TILLET, Barbara B. Authority control on the web [documento HTML]. En: *Bicentennial Conference on Bibliographic Control for the New Millennium: Confronting the Challenges of Networked Resources and the Web (2000)*.

Bibliografía

Washington). Sao Paulo: Universidade de Campinas, 10 de agosto de 2000, rev. 10 de agosto de 2001. Disponible en: http://www.unicamp.br/bc/tillett_paper.html

TILLMAN, Hope N. *Evaluating Quality on the Net* [documento HTML]. Hopetillman.com, 1995, rev. 19 de marzo de 2001. Disponible en: <http://www.hopetillman.com/findqual.html>

TOMINAGA, Mana. From the Expo Floor: Atomz [documento HTML]. *Webreview*, 7 september 2001. Disponible en: http://www.webreview.com/web2001/product_news/index02.shtml

TOMS, Eleaine G. Serendipitous Information Retrieval [documento PDF]. En: *DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries (2000. Zurich)*. ERCIM Workshop Proceedings, N° 01/W001 [2001]. Disponible en: http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf

TOZER, Guy. *Metadata Management for Information Control and Business Success*. Boston, London: Artech House, 1999.

TRIPPE, Bill. Taxonomies & Topic Maps: Categorization: Steps Forward. *EContent*, 2001, vol. 24, n° 6, p. 44-49.

TSA CONSULTING GROUP. *Automatic Generation of Metadata Based on Concepts Within the Document* [meta-data whitepaper] [documento HTML]. Klarity.com, 4 de Julio de 2000. Disponible en: http://www.klarity.com.au/whitepaper/meta-data_white_paper-06.htm#P70_7248

TURNER, Thomas P., Lise Brackbill. Rising to the Top: Evaluating the Use of the HTML META Tag to Improve Retrieval of World Wide Web Documents through Internet Search Engines. *Library Resources & Technical services*, 1998, vol. 42, n° 4, p. 258-271.

V

VAQUERO, Antonio. La corrupción del lenguaje por la informática [documento HTML]. *El Mundo*, 17 de abril de 1997. Disponible en: <http://www.el-mundo.es/1997/04/17/opinion/17N0017.html>

VAQUERO, Antonio. La lengua española en el contexto informático [documento HTML]. *Novática*, Julio-agosto 1999, n° 140. Disponible en: <http://www.ati.es/novatica/1999/140/antvaq140.html>

VARJONEN, Ville. DC/RDF/XML/XSLT/HTML, a Poor Man's Metadata Database [documento HTML]. Electronic publications of University of Oulu, 2001. Disponible en: <http://herkules.oulu.fi/vili/presentations/raju2001>

Bibliografía

VELLUCCI, Sherry L. Metadata. *Annual Review of Information Science and Technology (ARIST)*, 1998, vol. 33, p. 187-222.

VELLUCCI, Sherry L. Metadata and Authority Control. *Library Resources & Technical Service*, 2000, vol. 44, nº 1, p. 33-43.

VELLUCCI, Sherry L. Options for Organizing Electronic Resources: the Coexistence of Metadata. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, nº 1, p. 14-17.

VERGEL, Hal. A Digital Library as a Portal [documento HTML]. *DL Pearls*, June 2001. Disponible en: http://www1.acm.org/dl/dlpearls/dlp_6-01.html

VIDAL BORDÉS, Francisco Javier, José Antonio Salvador Oliván. La implementación de metadatos y Dublin Core en sedes y páginas web de bibliotecas y centros de documentación de universidades y centros de investigación de la Red Iris. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao). La gestión del conocimiento: Retos y soluciones de los profesionales de la información*. [Bilbao]: Universidad del País Vasco, 2000, p. 197-209.

VISUAL Resources Association Data Standards Committee. *VRA Core Categories: Version 3.0* [documento HTML]. Harvard University, 24 de julio de 2000. Disponible en: <http://www.gsd.harvard.edu/%7Estaffaw3/vra/vracore3.htm>

VIZINE-GOETZ, Diane. From Book Classification To Knowledge Organization: Improving Internet Resource Description and Discovery. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, nº 1, p. 24-27.

VOCABULARY ML: Metacode Strawman DTD [documento RTF]. NKOS, University of Kent, 11 de octubre de 2000. Disponible en: <http://nkos.slis.kent.edu/VOCML-1.DOC>

W

WAND, Patricia A. *Report from the OCLC Users Council, February 11 – 13, 2001* [documento PDF]. *CAPCON ALERT Special Edition*, March 2001, nº 65. <http://www.capcon.net/alertArchives/March%202001.pdf>

WARD, Nigel. Experiences with Distributed Searching [documento PDF]. En: *VALA Biennial Conference and Exhibition (10. 2000. Melbourne)*. Disponible en: <http://www.vala.org.au/vala2000/2000pdf/Ward.PDF>

WARD, Nigel and Andrew Wood. Emerging Technologies for Networked Information Discovery: Beyond Z39.50 [documento HTML]. En: *Resource Discovery Unit*. Brisbane: University of Queensland, Distributed Systems

Bibliografía

Technologies, rev. 15 de diciembre de 1997. Disponible en: <http://www.dstc.edu.au/RDU/reports/VALA98>

WAKE, Susannah, Dennis Nicholson. HILT-High Level Thesaurus Project [documento HTML]. *D-Lib Magazine*, September 2001, vol. 7, nº 9. Disponible en: <http://www.dlib.org/dlib/september01/wake/09wake.html>

WARWICK, Claire and Elliott Pritchard. 'Hyped' Text Markup Language. XML and the Future of Web Markup. *Aslib Proceedings*, May 2000, vol. 52, nº 5, p. 174-184.

WATSTEIN, Sarah B., Pascal V. Calarco, James S. Ghaphery. Digital library: Keywords. *Reference Services Review*, 1999, vol. 27, nº 4, p. 344-352.

WAUGH, Andrew. Specifying Metadata Standards for Metadata Tool Configuration [documento HTML]. En: *International World Wide Web Conference (7. 1998. Brisbane)*, rev. 9 de septiembre de 1998. Disponible en: <http://www7.scu.edu.au/programme/fullpapers/1913/com1913.htm>

WEBBER [documento HTML]. RedIris, rev. 15 de febrero de 2001. Disponible en: <http://www.rediris.es/app/webber>

WEBER, Mary Beth. Factors to be Considered in the Selection and Cataloging of Internet Resources. *Library Hi Tech*, 1999, vol. 17, nº 3, p. 298-303.

WEBOPEDIA: *Definition and Links* [documento HTML]. INT Media Group Incorporated, 2000. Disponible en: <http://webopedia.internet.com>

WEI CHOO, Chun, Brian Detlor and Don Turnbull. *Web Work: Information Seeking and Knowledge Work on the World Wide Web*. Dordrecht, Boston, London: Kluwer Academic Publishers, 2000.

WEIBEL, Stuart. The Dublin Core: A Simple Content Description Model for Electronic Resources. *Bulletin of the American Society for Information Science (Organizing Internet: Metadata and the Web)*, October/November 1997, vol. 24, nº 1, p. 9-11.

WEIBEL, Stuart. *A Proposed Convention for Embedding Metadata in HTML* [documento HTML]. En: *Distributed indexing/searching workshop (1996. Cambridge)*. World Wide Web Consortium, 2 de junio de 1996, rev. 20 de junio de 1996. Disponible en: <http://www.w3.org/Search/9605-Indexing-Workshop/ReportOutcomes/S6Group2.html>; también disponible en: <http://staff.oclc.org/~weibel/html-meta.html>

WEIBEL, Stuart. Metadata: the Foundations of Resource Description [documento HTML]. *D-Lib Magazine*, July/August 1995, vol. 1, nº 7/8. Disponible en: <http://www.dlib.org/dlib/July95/07weibel.html>

Bibliografía

WEIBEL, Stuart, Eric Miller. Image Description on the Internet: a Summary of the CNI/OCLC Image Metadata Workshop, September 24-25, 1996, Dublin, Ohio [documento HTML]. *D-Lib Magazine*, January 1997, vol. 3, n° 1. Disponible en: <http://www.dlib.org/dlib/january97/oclc/01weibel.html>

WEIBEL, Stuart et al. *Dublin Core Metadata for Resource Discovery: RFC 2413* [documento TXT]. IETF, Network Working Group, September 1998. Disponible en: <http://www.ietf.org/rfc/rfc2413.txt>

WEIBEL, Stuart, et al. *OCLC/NCSA Metadata Workshop Report* [documento HTML]. OCLC, marzo de 1995. Disponible en: http://www.oclc.org:5046/conferences/metadata/dublin_core_report.html (consultado el 15 de junio de 1998).

WEIBEL, Stuart, Juha Hakala. DC-5. The Helsinki Metadata Workshop: A Report on the Workshop and Subsequent Developments [documento HTML]. *D-Lib Magazine*, February 1998, vol. 4, n° 2. Disponible en: <http://www.dlib.org/dlib/february98/02weibel.html>

WEIBEL, Stuart, Renato Ianella, Warwick Catho. The 4th Dublin Core Metadata Workshop Report [documento HTML]. *D-Lib Magazine*, June 1997, vol. 3, n° 6. Disponible en: <http://www.dlib.org/dlib/june97/metadata/06weibel.html>

WEIBEL, Stuart, Traugott Koch. The Dublin Core Metadata Initiative: Mission, Current Activities, and Future Directions [documento HTML]. *D-Lib Magazine* December 2000, vol. 6, n° 12, Disponible en: <http://www.dlib.org/dlib/december00/weibel/12weibel.html>

WEINHEIMER, James L. How to Keep the Practice of Librarianship Relevant in the Age of the Internet. *Vine (Theme issue: Metadata. Part 1)*, 1999, n° 116, p. 14-37.

WHAT is a Bridge or Doorway Page? [documento HTML]. Search Engine Watch, 2 de junio de 2000. Disponible en: <http://searchenginewatch.internet.com/webmasters/bridge.html>

WHAT is RDF? Seminar [documento HTML]. Brian Kelly, ed. Bath: University, UKOLN, 12 de mayo de 1998. Disponible en: <http://www.ukoln.ac.uk/web-focus/events/seminars/what-is-rdf-may1998>

WHEATLEY, A. and C. J. Armstrong. Metadata, Recall, and Abstracts: Can Abstracts Ever be Reliable Indicators of Document Value? *Aslib Proceedings*, September 1997, vol. 49, n° 8, p. 206-213.

WILKINSON, Gene L., Kevin M. Oliver, Lisa T. Bennett. *Evaluating the Quality of Internet Information Sources* [documento HTML]. Georgia: Department of Instructional Technology, University of Georgia, 20 de mayo de 1997. Disponible en: <http://itech18.coe.uga.edu/faculty/gwilkinson/webeval.html>

Bibliografía

WINKWORTH, Ian. HyLife: the Hybrid Library of the Future. *The Electronic Library*, January 1999, vol. 17, nº 1, p. 17-19

WOODLEY, Mary S. Glossary [documento HTML]. DCMI, 12 de abril de 2001, rev. 24 de abril de 2001. Disponible en: <http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml> También disponible en: <http://library.csun.edu/mwoodley/dublincoreglossary.html>

WORKING GROUP ON DUBLIN CORE IN MULTIPLE LANGUAGES. *DC-Multilingual: Plans for a Distributed Registry of Dublin Core in Multiple Languages* [documento HTML]. Asian Institute of Technology, School of Advanced Technologies, 14 de julio de 1998. Disponible en: <http://www.cs.ait.ac.th/~tbaker/DC-Multilingual/Dcml.position> (consultado el 17 de julio de 1998).

WORLD STANDARDS SERVICE NETWORK. *General Info on Standardization* [documento HTML]. WSSN, 15 de noviembre de 1999. Disponible en: http://www.wssn.net/WSSN/gen_inf.htm

WORLD WIDE WEB CONSORTIUM².

— *The Cambridge Communiqué* [documento HTML]. Ralph R. Swick, Henry S. Thompson, eds. W3C, 20 de septiembre de 1999. Disponible en: <http://www.w3.org/1999/08/27-communique-19990920>

— *Extensible Markup Language (XML): W3C Working Draft 14-Nov-96* [documento HTML]. Tim Bray, C. M. Sperberg-McQueen, eds. W3C, 14 de noviembre de 1996. Disponible en: <http://www.w3.org/TR/WD-xml-961114.html>

— *Extensible Markup Language (XML) 1.0. W3C Recommendation 10-February-1998* [documento HTML]. Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, eds. W3C, 10 de febrero de 1998, rev. 9 de mayo de 2000. Disponible en: <http://www.w3.org/TR/1998/REC-xml-19980210>

— *Extensible Markup Language (XML) 1.0 (Second Edition). W3C Recommendation 6 October 2000* [documento HTML]. Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, eds. W3C, 5 de octubre de 2000. Disponible en: <http://www.w3.org/TR/2000/REC-xml-20001006>

² Se recogen a continuación todas las normas, recomendaciones y trabajos del Consorcio Web citados y/o consultados a lo largo de esta investigación cuya responsabilidad recae en el propio W3C como entidad, sin que ello obste para que se encuentren consignados en esta bibliografía otros documentos específicos que tienen un autor personal determinado aunque se hayan producido en el marco de la actividad del Consorcio.

Bibliografía

— *HTML 3.2 Reference Specification: W3C Recommendation, 14-Jan-1997* [documento HTML]. Dave Raggett. W3C, 14 de enero de 1997. Disponible en: <http://www.w3.org/TR/REC-html32>

— *HTML 4.0 Specification: W3C Recommendation, revised on 24-Apr-1998* [documento HTML]. Dave Raggets, Arnaud Le Hors, Ian Jacobs, eds. W3C, 24 de abril de 1998. Disponible en: <http://www.w3.org/TR/1998/REC-html40-19980424>

— *HTML 4.01 Specification W3C Recommendation 24 December 1999* [documento HTML]. Dave Raggets, Arnaud Le Hors, Ian Jacobs, eds. W3C, 24 December 1999. Disponible en: <http://www.w3.org/TR/1999/REC-html401-19991224>

— *Metadata: Activity Statement* [documento HTML]. W3C, rev. 10 de octubre de 2000. Disponible en: <http://www.w3.org/Metadata/Activity.html>

— *RDF Model Theory. W3C Working Draft 25 September 2001* [documento XHTML]. Patrick Hayes, ed. W3C, 25 de septiembre de 2001. Disponible en: <http://www.w3.org/TR/2001/WD-rdf-mt-20010925>

— *RDF Test Cases. W3C Working Draft 15 November 2001* [documento XHTML] Art Barstow, Dave Beckett, eds. W3C, 15 de noviembre de 2001. Disponible en: <http://www.w3.org/TR/2001/WD-rdf-testcases-20011115>

— *Refactoring RDF/XML Syntax. W3C Working Draft 06 September 2001* [documento HTML]. W3C, 6 de septiembre de 2001. Disponible en: <http://www.w3.org/TR/2001/WD-rdf-syntax-grammar-20010906>

— *Resource Description Framework (RDF) Model and Syntax Specification. W3C Recommendation 22 February 1999* [documento HTML]. Ora Lassila and Ralph R. Swick, eds. W3C, 24 de febrero de 1999. Disponible en: <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222>

Traducción de Eva Méndez en: <http://www.sidar.org/traduc/rdfesp.htm> y en: <http://www.bib.uc3m.es/~mendez/RDF/syntax/REC-rdf-syntax.htm>

— *Resource Description Framework (RDF) Schema Specification 1.0 W3C Candidate Recommendation 27 March 2000* [documento HTML]. Dan Brickley R.V. Guha, eds. W3C, 27 de marzo de 2000. Disponible en: <http://www.w3.org/TR/2000/CR-rdf-schema-20000327>

Traducción de Eva Méndez en: <http://www.sidar.org/traduc/rdfsch.htm> y en: <http://www.bib.uc3m.es/~mendez/RDF/schema/CR-rdf-schema.htm>

— *Web Content Accessibility Guidelines 1.0: W3C Recommendation 5-May-1999* [documento HTML]. Wendy Chisholm, Gregg Vanderheiden, Ian Jacobs, eds. W3C, 5 de mayo de 1999. Disponible en: <http://www.w3.org/TR/1999/WAI-WEBCONTENT-19990505>

Bibliografía

- *XHTML™ 1.0: The Extensible HyperText Markup Language: A Reformulation of HTML 4 in XML 1.0. W3C Recommendation 26 January 2000* [documento XHTML]. W3C, 26 de enero de 2000. Disponible en: <http://www.w3.org/TR/2000/REC-xhtml1-20000126>
- *XHTML™ 1.0: The Extensible HyperText Markup Language: A Reformulation of HTML 4 in XML 1.0 (Second Edition). W3C Working Draft 4 October 2001* [documento XHTML]. W3C, 4 de octubre de 2001. Disponible en: <http://www.w3.org/TR/2001/WD-xhtml1-20011004>
- *XML Information Set. W3C Recommendation 24 October 2001* [documento XHTML]. Jonh Cowan, Richard Tobin, eds. W3C, 24 de octubre de 2001. Disponible en: <http://www.w3.org/TR/2001/REC-xml-infoset-20011024>
- *XML Linking Language (XLink) Version 1.0. W3C Recommendation 27 June 2001* [documento HTML]. Steve DeRose, Eve Maler, David Orchard, eds. W3C, 27 de junio de 2001. Disponible en: <http://www.w3.org/TR/2001/REC-xlink-20010627>
- *XML Schema Part 0: Primer. W3C Recommendation, 2 May 2001* [documento HTML]. David C. Fallside, ed. W3C, 2 de mayo de 2001. Disponible en: <http://www.w3.org/TR/2001/REC-xmlschema-0-20010502>

X

XML la penúltima revolución de Internet [documento HTML]. *Infosociedad*, 2 de septiembre de 1998. Disponible en: http://www-5.ibm.com/es/infosociedad/experiencias/tecnologia_9808.htm

XU, Amanda. Accessing information on the Internet: feasibility study of USMARC format and AACR2 [documento HTML]. En: *OCLC Internet Cataloging Colloquium (1996. San Antonio)*. OCLC, 19 de enero de 1996. Disponible en: <http://www.oclc.org/oclc/man/colloq/xu.htm>

XU, Amanda. *Metadata Conversion and the Library OPAC* [documento HTML]. Massachusetts: Institute of Technology, rev. 30 de enero de 1998; The Haworth Press, 1997. Disponible en: <http://web.mit.edu/waynej/www/xu.html> (consultado el 25 de junio de 2001).

Y

YEATES, Robin. Metadata: Introduction. *Vine (Theme issue: Metadata. Part 1)*, 1999, nº 116, p. 3-5.

Bibliografía

YI, Jeonghee, Neel Sundaresan, Anita Huang. Using Metadata to Enhance Web Information Gathering. En: *The World Wide Web and Databases. International Workshop WebDB (3. 2000. Dallas)*. Dan Suciu; Gottfried Vossen, eds. Berlin, etc.: Springer, 2001, p. 38-57.

Z

Z39.50 *Profile for Access to Digital Collections*. Washington, DC.: Library of Congress, 3 de mayo de 1996, rev. 13 de enero de 1997. Disponible en: <http://lcweb.loc.gov/z3950/agency/profiles/collections.html>

ZANETE, Nelson Henrique. *A Contribution to Brian McBride's Approach* [documento HTML]. Porto Alegre: FACCAR, Universidade Federal de Rio Grande do Sul, 6 de julio de 2001. Disponible en: <http://www.faccar.com.br/zanete/rdfdb.htm>

ZANETE, Nelson Henrique. *Introdução ao RDF: Resource Description Framework* [documento HTML]. Porto Alegre: FACCAR, Universidade Federal de Rio Grande do Sul, 13 de febrero de 2001. Disponible en: <http://www.faccar.com.br/zanete/zaneteRDF.htm>

ZELNICK, Nate. Getting Deep into Metadata [documento HTML]. *Webdeveloper.com*, 24 de junio de 1998. Disponible en: http://www.webdeveloper.com/xml/xml_061298.html (Part II) y http://www.webdeveloper.com/xml/xml_061298_part2.html (Part II)

ZENG, Marcia Lei. Metadata Elements for Object Description and Representation: A Case Report from a Digitized Historical Fashion Collection Project. *Journal of the American Society for Information Science (Special Topic Issue: Integrating Multiple Overlapping Metadata Standards)*, 1999, vol. 50, nº 13, p. 1193-1208.

ZHAO, Ben. Y., Anthony D. Joseph. *Xset: A Lightweight Database for Internet Applications* [documento PDF]. Berkeley: University of California, Computer Science Division, rev. 5 de junio de 2000. Disponible en: <http://www.cs.berkeley.edu/~ravenben/xset/html/xset-saint.pdf>

ZICK, Laura. The Work of Information Mediators: A Comparison of Librarians and Intelligent Software Agents [documento HTML]. *First Monday*, May 2000, vol. 5, nº 5. Disponible en: http://www.firstmonday.dk/issues/issue5_5/zick/index.html

ZUBELDIA LAUZURICA, M^a Isabel. Cooperación frente al caos en Internet: CORC, una propuesta de trabajo. En: *Jornadas Españolas de Documentación (7. 2000. Bilbao)*. La gestión del conocimiento: Retos y soluciones de los profesionales de la información. [Bilbao]: Universidad del País Vasco, 2000, p. 597-606.

**GLOSARIO EXPLICATIVO DE LA TERMINOLOGÍA
UTILIZADA**

Glosario

Application profile

El hecho de incluir este concepto en el glosario no se debe a la dificultad en su traducción (perfil de aplicación), sino a la vaguedad de ésta en nuestro idioma. Se trata de un concepto muy concreto introducido por Rachel Heery en el congreso del Dublin Core de Ottawa (2000), desarrollado posteriormente en consonancia con RDF.

Un perfil de aplicación es un esquema (*schema*) de metadatos que consiste en un conjunto de elementos trazados o seleccionados a partir de uno o varios *namespaces*, combinados por los desarrolladores de un determinado formato (como el DC) y optimizados para una aplicación particular¹, como por ejemplo las bibliotecas (*library application profile*). Un perfil de aplicación es pues, un conjunto de entidades que sirven para declarar qué elementos, de qué espacios de nombre (*namespaces*) subyacen a un esquema (*schema*) local utilizado en un proyecto de aplicación particular. Tratan de reutilizar la semántica de un *namespace* de propósito general, con un propósito particular adaptándolo, al incluir otros elementos, a las necesidades específicas de la información que se quieren describir y por tanto, recuperar.

Vid. Namespace, Schema/Scheme.

Best practice

Esta expresión anglosajona quiere decir literalmente "práctica mejor", sin embargo su uso en inglés y en la bibliografía relacionada con los metadatos lleva asociado un significado semejante a "manual de procedimiento" o conjunto de experiencias probadas en el ejercicio de una función.

La mejor conceptualización de este término viene de la mano del DCMI, donde se define como la *guía y documentación para describir y normalizar el uso de elementos de metadatos que mejor soportan las necesidades de una comunidad*². Así, cada proyecto de aplicación de metadatos elabora un

¹ Rachel Heery, Manjula Patel. *Op. cit.*, <http://www.ariadne.ac.uk/issue25/app-profiles>

² Mary S. Woodley. *Glossary. Op. cit.* <http://dublincore.org/documents/2001/04/12/usaguide/glossary.shtml#B>

documento de estas características donde se recoge su propia interpretación *ad hoc* de un modelo de metainformación concreto.

Broker

El significado general de *broker* es "agente o corredor de bolsa". Sin embargo, en este trabajo, dicho vocablo encierra una gran complejidad de utilización en español debido, por un lado, al significado que ha adquirido en el área de la Biblioteconomía y Documentación (*vid. brokering*), y por otro, a su uso como componente de software en los sistemas de indización de Internet (sentido que reflejamos ahora).

En el ámbito de la Documentación, el uso más generalizado, recogido en algunos diccionarios especializados³ para este término en inglés es el de *information broker* o *information brokering*, que trataremos a continuación. Por otra parte, en el contexto de la recuperación de información en Internet, un *broker* es un software mediador o intermediario que sirve para relacionar dos aplicaciones. En este sentido lo reconocen Dempsey, Russel y Heery⁴, al hablar de *broker* en el contexto de los metadatos, como *un nivel de software* que posibilita el acceso entre los usuarios y los recursos disponibles. Este significado de *broker* —al igual que en el caso de *harvest/gatherer* (*Vid. infr.*)— podría derivarse de la denominación que recibe uno de los subsistemas del software Harvest. En este conjunto de herramientas, el *broker* recupera información indizada de alguno de los *gatherers* y sirve para suprimir información duplicada, indizar de forma creciente la información recopilada y proporcionar una interfaz de consulta Web⁵. Sin embargo, el *broker* no es exclusivo de Harvest sino que forma parte de otros sistemas de recuperación de información en Internet como el modelo de programación CORBA, que basa su nombre en uno de los componentes, el *Object Request Broker (ORB)* que actúa

³ Antonio Lozano Palacios. *Op. cit.*, p. 51-52.

⁴ Lorcan Dempsey, Rosemary Russell, Rachel Heery. In at the Shallow End: Metadata and Cross-domain Resource Discovery. En: *Discovering Online Resources across the Humanities...* *Op. cit.*, p. 63; http://ahds.ac.uk/public/metadata/disc_07.html (en la versión digital del libro).

⁵ Para más detalles sobre el funcionamiento de Harvest, *Vid.* Darren R. Hardy, Michael F. Schwartz and Duane Wessels. *Harvest User's Manual (Version 1.4 patchlevel 2)* [documento HTML]. Harvest, 31 de enero de 1996, rev. 22 de abril de 1996. Disponible en: <http://www.tardis.ed.ac.uk/harvest/docs/old-manual/user-manual.html> (consultado el 18 de marzo de 2001).

como intermediario entre el cliente y el servidor. En general y dependiendo de su riqueza funcional, este nivel de software puede ocultar la heterogeneidad de los recursos accesibles, puede proporcionar soporte para la navegación y selección y puede facilitar que los usuarios interactúen de diversas formas con los recursos.

Cfr. Brokering (information brokering).

Brokering

Como anticipábamos en la voz anterior, tal y como se entiende en general en el mundo de la Información y Documentación, un *broker* de información es un organismo comercial, a veces una persona, que se dedica a localizar, recuperar y ofrecer información sobre cualquier materia, algo así como un gestor de información por cuenta propia. Sin embargo, no existe un como verbo en inglés de la misma raíz, por ello, en el caso de *information brokering* tratamos de cargar al participio presente con la acción que se puede deducir del sustantivo, esto es, el proceso de recuperación y disposición de información.

Levine⁶ define el *brokering* de información como *el negocio de comparar y vender información como un activo*, se trata de un concepto emanado de la *Société Française de Radiophonie* en 1935, y consistía en trascender la función de la biblioteca de distribución de la información más allá de sus canales formales. En la Era electrónica, con el crecimiento de la World Wide Web, este concepto ha tomado un nuevo cariz, al sobrestimarse la necesidad de recuperar conocimiento útil. No obstante, en esta investigación, entendemos *information brokering* como *una arquitectura que guía en la creación y gestión (brokering) de sistemas y soluciones*, y como *un proceso de mediación entre los usuarios y los proveedores de información que plantea una solución clave para afrontar el problema de la sobrecarga de información*⁷. Se trata pues, de un proceso de mediación para salvar la heterogeneidad y la inflación de información y hacerla accesible en un contexto global. El *brokering* se da en tres niveles: representación, contenido de la información y vocabulario, y tiene como

⁶ Marilyn M. Levine. A Brief History on Information Brokering [documento HTML]. *Bulletin ASIS*, February/March 1995, vol. 21, nº 3. Disponible en: <http://www.asis.org/Bulletin/Feb-95/levine.html> (consultado el 4 de febrero de 2001).

⁷ Vid. 3.4. Aplicación de los metadatos. Virpul Kashyap, Amit Sheth. *Information Brokering...*, *Op. cit.*, p. 6, 209.

resultado una reducción del volumen y de los datos que deben tratarse, donde los metadatos toman una preeminencia fundamental, en el sentido de que proporcionan una nueva arquitectura de la información basada en la escalabilidad, la extensibilidad y la adaptabilidad.

Cfr. Broker.

Browser

Vid. Browsing.

Browsing

Aunque originalmente el verbo "*to browse*" en inglés significa hojear un libro, hacer una lectura ligera, con el desarrollo de Internet, el participio presente de este verbo se utiliza de forma genérica para denominar la actividad de leer páginas en Internet pasando de una a otra siguiendo hiperenlaces, esto es, lo que se conoce en Internet con el término metafórico de "navegar" o "navegación". En consonancia con esta actividad, el software agente de usuario, que permite seguir el hipertexto codificado en HTML, recibe la denominación de *browser*⁸.

En esta tesis utilizamos el término con varios matices distinguiendo entre el *browsing* "convencional", hipertextual, que acabamos de definir, para el que utilizamos la palabra en inglés directamente obviando el uso de su traducción (navegación, navegar). Utilizamos el término en español (navegación) para referirnos a un *browsing* dirigido y estructurado, o sea, para denominar la técnica de recuperación de información basada en clasificaciones jerárquicas⁹ (navegación dirigida).

⁸ El diccionario especializado en términos de informática, What?is. <<http://whatis.techtarget.com>> da una definición concisa, a la par que explícita del término *browser*: Un *browser* (o navegador, en español) es una aplicación informática que proporciona una forma de mirar e interactuar con la información de la World Wide Web. Más información sobre este tipo de aplicaciones en: <http://whatis.techtarget.com/wsearchResults/1,290214,sid9,00.html?query=browser>

⁹ *Vid. 7.1.1.*

Clearinghouse/s

Este término significa "banco de liquidación" o "cámara de compensación" en inglés convencional. Aplicado a la información electrónica se refiere a un *centro de proceso de información. Organismo cuya tarea es la de recoger, conservar, difundir y poner a disposición documentos e información sobre estudios e investigaciones completas, propuestas o en curso, así como enviar a los usuarios a otras fuentes de información*¹⁰.

En relación con la metainformación, un *clearinghouse* es un tipo especial de *subject gateway* (*Vid. infr.*), es decir, un sistema de información digital temático, normalmente sobre información geoespacial, organizado según un esquema de metadatos.

Se trata de un concepto creado en 1994 en Estados Unidos durante la política de la *National Information Infrastructure* (NII) del gobierno Clinton, concretamente a través de la orden ejecutiva 12906 donde se define además la Infraestructura Nacional de Datos Espaciales (NSDI). Según dicha orden ejecutiva, un *clearinghouse* es *una red distribuida de productores, gestores y usuarios de datos geoespaciales, enlazados electrónicamente*¹¹; Posteriormente se generalizaría este término para denominar cualquier *servicio distribuido para localizar metadatos de datos geoespaciales que permite realizar búsquedas en uno o varios nodos o servidores registrados al clearinghouse de Internet*¹² [...].

La ventaja principal de estos sistemas de búsqueda finitos o bibliotecas digitales temáticas, al igual que otros basados en metadatos, es que el *clearinghouse* no permite solamente la búsqueda textual sino también en bases de datos constituidas bajo el estándar FGDC o a través de regiones o

¹⁰ Antonio Lozano Palacios. *Op. cit.*, p. 68.

¹¹ Dicha orden dice textualmente: *The "National Geospatial Data Clearinghouse" means a distributed network of geospatial data producers, managers, and users linked electronically.* Willam Clinton. *Executive Order 12906*. Published in the April 13, 1994, edition of the Federal Register, Volume 59, Number 71, pp. 17671-17674 [documento HTML]. FGDC, [1994]. Disponible en: <http://www.fgdc.gov/publications/documents/geninfo/execord.html> (consultado el 2 de junio de 2001).

¹² *Clearinghouse de datos geográficos de El Salvador: preguntas frecuentes* [documento HTML]. San Salvador: Ministerio de Medio Ambiente, rev. 23 de abril de 2001. Disponible en: <http://www.marn.gob.sv/clearinghouse/menufoot/preg.htm> (consultado el 2 de junio de 2001).

coordenadas geográficas (en el caso de que su temática sea la geoespacial). Las principales características de estos sistemas de información temática son la coherencia del modelo de metadatos, el nivel de desarrollo y el software específico (fundamentalmente MetaLite) para todos los *clearinghouses* del mundo¹³.

En el marco de la Documentación, no existe un consenso en la traducción, ni siquiera una traducción específica de este término en español. Por ello, para referirnos al concepto que evoca, hemos utilizado directamente el término en inglés o bien una explicación idiomática en nuestra lengua que evoque su significado, como "sistemas de información geoespacial".

Vid. Subject gateway/Information gateway.

Crosswalks/Crossroads

Paso de peatones y cruce de carreteras, respectivamente, son los significados de la traducción literal de estos términos, cuya utilización en un trabajo de estas características resultaría un tanto incongruente —igual que ocurre con el término *harvesting* como veremos— y por ello utilizamos de forma genérica el término en inglés. Preferentemente usamos *crosswalks* porque es el más aceptado en la bibliografía sobre el tema. El término *crossroads* se acuñó en el seno del proyecto ROADS¹⁴ (*Resource Organisation And Discovery in Subject-based Services*) donde se denomina "*crossROADS*" a su trabajo de integración de servicios de información e interoperabilidad, aprovechando el significado semejante y la terminación final idéntica al acrónimo del proyecto (ROADS).

Pese a lo singular de su traducción, estos términos llevan aparejado un concepto fundamental en las investigaciones sobre metadatos, que alude a las correspondencias o conversiones (mapeo, *vid. mapping*) que se tienen que realizar en los sistemas de información que utilizan distintos modelos de

¹³ Esta coherencia y nivel de desarrollo la apuntamos dentro de las conclusiones de esta investigación, refiriéndonos entonces a los metadatos geoespaciales. Justo la coherencia y estandarización de su metainformación, trasciende a los sistemas de información (*clearinghouses*) basados en tales metadatos. *Cfr.* Capítulo 10, conclusión nº 25.

¹⁴ Sobre este proyecto, *Vid.* Rachel Heery, Andy Powel and Michael Day. Metadata: CrossROADS... *Op. cit.*, <http://www.ariadne.ac.uk/issue14/metadat>

metadatos en su descripción. Es una forma metafórica de hablar de la interoperabilidad entre sistemas, entendiendo ésta por la capacidad que tienen dos sistemas, de diferente tipo y estructura de datos, de funcionar conjuntamente y de comunicarse. De una forma técnicamente más específica, Majula Patel¹⁵ de la UKOLN, define *crosswalk* como un 'mapeo' desde los elementos de un namespace (Vid. *Infr.*) a los elementos de otro namespace, donde las *crosswalks* son la representación visual o los "mapas que representan las relaciones entre dichos elementos". A la complejidad de este concepto, se une además el hecho —reconocido por St Pierre y LaPlant¹⁶— de que no existe una terminología estándar para denominar esos bits de información, que conlleva que en diferentes sistemas se refieran a ellos como *campos* (*fields*), etiquetas (*labels*), marcas (*tabs*), o identificadores (*identifiers*).

Vid. *Interoperability, Mapping, Namespace*.

Datawarehouse

Aunque la traducción es sencilla, "almacén de datos", si se utiliza en español pierde carga semántica ya que se produce una simplificación inminente del contenido de esta expresión en inglés, que se utiliza profusamente en el ámbito de la gestión de la información empresarial. A veces, también se emplea para designar lo mismo, gracias al influjo de IBM, la expresión *information warehouse*. Un *data/information warehouse* es un depósito central para todos los datos, o una parte muy significativa de los datos que se acumulan en varios sistemas de negocio dentro de una empresa.

De una forma más sencilla y siguiendo a Tozer¹⁷, se puede decir que el concepto de *data warehouse* se creó para definir un modelo para el suministro de información dentro de grandes y complejas organizaciones. Se trata de un paradigma de organización y gestión de la información empresarial que, en

¹⁵ Glossary [documento HTML]. En: *Schemas: Forum for Metadata Schema Implementers*. Maintained by Manjula Patel. Schemas, 2 de agosto de 2000. Disponible en: <http://www.schemas-forum.org/related/glossary.html#c> (consultado el 10 de abril de 2001).

¹⁶ Margaret St. Pierre, William P. LaPlant, Jr. Issues in Crosswalking, Content Metadata Standards [documento HTML]. *NISO Standards*. Bethesda: NISO, 15 de octubre de 1998. Disponible en: <http://www.niso.org/crswalk.html> (consultado el 5 de octubre de 2000).

¹⁷ Guy Tozer. *Op. cit.*, p. 108-109, 296.

términos prácticos, se suele convertir en una gran base de datos centralizada que abarca la gestión de la información de toda la empresa y que normalmente se nutre de aplicaciones tipo OLAP (*Online Analytical Processing*).

Aunque el contexto informativo del *data warehouse* queda fuera de los objetivos de esta investigación, utilizamos esta expresión en el sentido de que cada vez más estos sistemas de almacenamiento y gestión de datos están tomando modelos de metainformación para controlar y recuperar la información; igualmente, muchos de los software comerciales de metadatos están dirigidos a este sector. Por otra parte, el hecho de que los entornos informativos finitos de las empresas adopten, de forma creciente, sistemas de metadatos para organizar, gestionar y recuperar su información, avala nuestra tesis de la funcionalidad de los metadatos en contextos controlables de información más que en el universo total de la World Wide Web.

Discovery

Aunque la traducción de esta palabra, en cualquier otro contexto, sea "descubrimiento", sólo la hemos traducido así en los casos en que era necesario poner de relieve el hallazgo de recursos de información frente a la mera localización de recursos previamente conocidos. En el resto de las ocasiones en que ha sido necesario hacer referencia a este concepto, hemos traducido *discovery* como búsqueda o recuperación de información en forma genérica, ya que el uso de la expresión *resource/information discovery* se asocia al proceso de búsqueda de información en Internet.

Ray Denenberg corrobora esta decisión de utilización del término al precisar que *information discovery significa localización de objetos de interés cuando el número de objetos entre los que se puede elegir es potencial y ampliamente distribuido*¹⁸, circunstancias éstas que confluyen en la búsqueda en Internet. Large, Tedd y Hartley¹⁹ también coadyuvan a matizar este significado a través de la explicación de la utilidad de los motores de búsqueda de Internet diciendo que estas herramientas de búsqueda resultan útiles para *localizar* recursos

¹⁸ ...*Information discovery means locating objects of interest when the population of object from which to choose is potentially widely distributed.* Ray Denenberg. *Op. cit.*, <http://lcweb.loc.gov/z3950/agency/papers/italy.html>

¹⁹ Andrew Large, Lucy A. Tedd and R. J. *Op. cit.*, p. 137.

conocidos en Internet, pero son menos satisfactorias para la recuperación de recursos (*resource discovery*).

DLO

En este caso, no es necesario esclarecer la cobertura de este acrónimo *Document Like Object* que ha sido extensamente explicada a lo largo de este trabajo (*Vid.* 3.2). Pero sí nos parece interesante justificar por qué hemos elegido la sigla en inglés para referirnos a los Documentos Como Objetos, en lugar de la correspondiente en español (DCO). Hemos utilizado DLO justamente porque al tratarse de un acrónimo, se puede pronunciar en español sin que sufra nuestro idioma (como ocurre con otras siglas y acrónimos, como radar) y sin perder el carácter internacionalizado y universal de la sigla.

Embedded

El verbo inglés *embed* significa incrustar, empotrar, encastrar. El adjetivo correspondiente, *embedded*, se utiliza en el ámbito de este trabajo para indicar que los metadatos están incluidos o incrustados en el propio código del documento electrónico. Se advierte no obstante, una tendencia generalizada a traducirlo al español a través una transliteración, cuando se habla de datos que aparecen "*embebidos*" en el documento.

En inglés, se utiliza el término *imbibe* (embeber), sólo con el significado de absorber, empapar o impregnar. Estamos pues, ante uno de los casos claros en que el inglés y el español tienen campos semánticos distintos y no tienen una correspondencia unívoca de significados: en inglés existen dos palabras (*embed*, *imbed*) de grafía muy parecida y alófonos desde el punto de vista de la pronunciación, que significan "incrustar", y una sola (*imbibe*) para decir embeber; mientras que la polisemia del castellano utiliza el mismo término (embeber) tanto en el sentido de absorber o empapar, como en el sentido de *contener, encerrar una cosa dentro de sí a otra*²⁰. Gracias a la polisemia del español, es adecuado el uso de embeber (y así lo utilizamos en la tesis) para designar los metadatos contenidos dentro del propio documento. Sin embargo,

²⁰ *Diccionario de la Academia Española. Op. cit.*, <http://buscon.rae.es/drae/drae.htm>, significado 3 de la voz "embeber".

desde nuestro punto de vista este uso se debe más al calco semántico que a la reflexión que hemos hecho aquí.

Format

Vid. Metadata [Format/Model/Registry/Schema/Standard].

Gateway

A pesar de que hoy en día se utiliza también el término *router* (traducido como direccionador, encaminador ó enrutador) en lugar de la definición original de *gateway*, en este trabajo lo hemos traducido como pasarela²¹, refiriéndonos a los programas o dispositivos de comunicaciones que transfieren datos entre redes que tienen funciones similares pero implantaciones diferentes. Una *gateway*, pasarela o intersección sirve para el intercambio de información a través de redes incompatibles que usan protocolos diferentes²². Este término es muy importante al hablar de la interoperabilidad entre sistemas de metadatos, y en ese sentido lo utilizamos en este trabajo. Sin embargo, también se usa para designar, de forma simplificada un conjunto de sitios de Internet que de alguna manera proporcionan acceso a otros recursos. Esos servicios están destinados a facilitar la recuperación de información sobre una materia concreta a un grupo de usuarios determinado; es lo que conocemos de forma más específica como *Subject gateways*.

Vid. Gateway page, Subject gateway/Information gateway.

Gateway page

La amplitud del término *gateway* ha dado lugar a expresiones como *gateway pages* o *subject gateways*. La expresión *gateway page* surge en el contexto del

²¹ Rafael Fernández Calvo. *Op. cit.*, <http://www.ati.es/PUBLICACIONES/novatica/glointv2.html>

²² *NetGlos - The Multilingual Glossary of Internet Terminology* [documento HTML]. World Wide Languages Institute, 15 de marzo de 1997, rev. 21 de febrero de 1998. Disponible en: <http://wwli.com/translation/netglos/glossary/glossary.html#G> (consultado el 17 de julio de 1998).

software Promotion-Pro²³, y es un concepto vinculado, como las metaetiquetas, a la promoción de la Web. Se trata de una página construida cuidadosamente, que acompañando a un sitio Web, sirve para alcanzar el mayor ranking en los motores de búsqueda.

Vid. Gateway.

Gather-Gathering

Vid. Harvest-Harvesting.

Granularity

Ante este término hemos utilizado, en la mayoría de los casos, la traducción libre "nivel de detalle" (en la descripción de recursos de información), ya que el calco al español "granularidad" es un poco desconcertante en el contexto de este trabajo y porque su traducción en un vocabulario especializado en documentación²⁴ como *falta de orden jerárquico*, tampoco es acertada en este ámbito.

Según Liora Alschuler²⁵, el concepto de "nivel de detalle" se puede explicar desde una perspectiva de estructuración lógica de la información (en su caso, en SGML), donde *granularity es la escala de una unidad de información típica; de tal forma que las marcas que identifican párrafos como unidad más pequeña, pueden constituir un nivel de "granularidad", mientras que las marcas que identifican partes del discurso pueden ser un nivel de detalle diferente y más refinado.*

²³ *Vid.* Capítulo 4, notas 124 y 125.

²⁴ Antonio Lozano Palacios. *Op. cit.*, p. 151.

²⁵ Glossary: Talking SGML. En: Liora Alschuler. *Op. cit.*, p. 396. En este caso Alschuler, se refiere a la información estructurada en *Standard Generalised Markup Language*, pero esa diferenciación de niveles de información se puede referir a cualquier ámbito de la información estructurada a través de metadatos.

Por otra parte, Rust²⁶ hace referencia al nivel de descripción que deben reflejar los metadatos con relación a una unidad lógica de información estructurada. Enuncia el principio de nivel de detalle (*granularity*) de la siguiente forma: *las entidades serán identificadas cuando necesiten ser identificadas*. Se refiere con ello a que los metadatos sólo deben tener un nivel de detalle descriptivo en un elemento (independientemente del formato que adopten) en el caso de que sea necesario para su identificación y diferenciación posterior. Al fin y al cabo, en el mundo de la catalogación tradicional siempre han existido políticas catalográficas sobre qué nivel de detalle se debe dar a la descripción, por ejemplo, si se vacían los artículos de una revista o se cataloga la revista como un solo documento. De igual modo, con el concepto de "nivel de detalle" (granularidad) se trata de determinar cuál es el mejor nivel de descripción de un recurso electrónico Web, por ejemplo, si un proyecto debe describirse a través de metadatos, como un único recurso, o si se deben detallar todas las secciones del sitio Web donde se describe (personas que participan en el proyecto, objetivos y detalles técnicos). Según esto, *cada formato de metadatos establecerá un nivel de detalle*²⁷.

Harvest-Harvesting /Gather-Gathering

El verbo inglés *to harvest*, en sentido general, en cualquier diccionario de inglés-español, significa: *cosechar, recoger, recolectar*²⁸, esquilmar, hacer la cosecha, y por ello su traducción literal no es oportuna en el contexto de esta investigación. Incluso en el *Vocabulario inglés/español - español/inglés para los estudios de biblio-documentación*²⁹ el sentido que se da a esta palabra es el mismo (cosechar, recoger). Esto nos ha llevado a recurrir a otras reflexiones para la utilización de este término en el contexto de la recuperación de información en Internet, que vayan más allá de la metáfora. Se debe a la

²⁶ Godfrey Rust. *Op. cit.*, <http://www.dlib.org/dlib/july98/rust/07rust.html#granularity>

²⁷ Cfr. Rachel Heery. A Review of Metadata... *Op. cit.*, <http://www.ukoln.ac.uk/metadata/review.html>

²⁸ Colin Smith. *Collins: Diccionario español-inglés inglés-español*. 5ª ed. Barcelona: Grijalbo Mondadori, 1997, p. 1129.

²⁹ Antonio Lozano Palacios. *Op. cit.*, p. 157.

existencia de un software de indización denominado Harvest³⁰, surgido a finales de 1993 dentro de la línea de trabajo del IRTF-RD (*Internet Research Task Force Research Group-Resource Discovery*) protagonizado por Michael F. Schwartz. A pesar de que este software puede ser el origen, en la bibliografía en inglés se utiliza el término *harvesting* para referirse de forma genérica a la indización de información basada en partes concretas de un documento marcado, normalmente, los metadatos. Concretamente, se ha definido el objetivo del *harvesting* como *preparar los metadatos para hacerlos útiles a los usuarios en la búsqueda de información*³¹, y este es el sentido con el que utilizamos el término en la tesis, cuya raíz han adoptado muchas herramientas de software que trabajan de esta forma, como por ejemplo la aplicación *MetaStar Harvester* de Blue Angel.

Siguiendo la metáfora pura de la cosecha para referirse a la recuperación de información, en el sistema de indización y recuperación distribuida Harvest se utiliza el término *gatherer* (recolectar) para denominar al subsistema que automatiza el proceso de extracción de información dentro de los ficheros de un mismo servidor. Sin embargo, y teniendo en cuenta que el término se usa también en contextos generales de recuperación de información, *gathering* es el proceso de recuperación u obtención, usado también para aludir al resultado de la búsqueda, en el sentido de la correlación de los términos cosechar (*harvest*) y recolectar (*gather*), que traducido al campo de la recuperación de información en Internet, se puede interpretar como indizar y recuperar.

³⁰ En realidad Harvest es un conjunto de utilidades que permiten recoger, extraer, organizar, buscar o reproducir información indizada en Internet. Para más información sobre este software y sistema de búsqueda. Mic Bowman, et al. *The Harvest Information Discovery and Access System* [documento HTML]. 2 de septiembre de 1996, rev. 10 de noviembre de 1997. Disponible en: <http://harvest.transarc.com> (consultado el 17 de julio de 1998). Para detalles concretos sobre el uso y funcionamiento de Harvest, Vid. Darren R. Hardy, Michael F. Schwartz and Duane Wessels, *Op. cit.*, <http://www.tardis.ed.ac.uk/harvest/docs/old-manual/user-manual.html>

A pesar de que en marzo del 2000 se anunció la versión 1.7.0 de Harvest, no es estable. Para obtener información sobre la versión estable (1.6.1), así como el propio software Vid. <http://www.tardis.ed.ac.uk/harvest> En España, se puede descargar la aplicación del servidor ftp de la Universidad de Oviedo: <http://www.si.uniovi.es/mirror/harvest-net> (consultado el 24 de agosto de 2000).

³¹ *Harvesting is the reverse operation of creation. The objective of harvesting is to prepare the metadata to make it useful for searchers.* Makx Dekkers. Harvesting: Breakout Session [presentación PPT]. En: *EC Metadata Workshop (1997. Luxembourg)*. Bath: UKOLN, rev. 12 de diciembre de 1997. Disponible en: <http://hosted.ukoln.ac.uk/ec/metadata-1997/harvest/harvest/harvest.ppt>, slide 1 (consultado el 27 de julio de 2000).

Cfr. *Harvest Control Lists*.

Harvest Control Lists (HCL)

La traducción de esta expresión sería algo así como "listas de control de indización". Pero consideramos que no se puede hacer una traducción tan genérica pues este concepto está asociado a los sistemas de información de las Agencias del gobierno australiano que utilizan en modelo AGLS³² (*Australian Government Locator Service*) para estructurar y recuperar su información, así como al software de visualización y edición de metadatos Metabrowser³³, que contempla la normativa del formato AGLS, donde estas listas de control de indización reciben la denominación de HCL, como hemos visto³⁴.

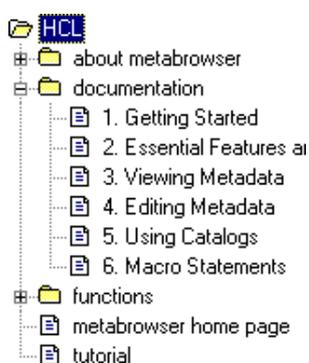


Fig. 69. HCL (Metabrowser)

En ambos contextos, se entiende por *Harvest Control List* una página índice, una especie de mapa del sitio, formado por una lista de enlaces a las páginas clave que tienen embebidos todos los metadatos AGLS, de tal forma que se simplifica la gestión y la indización de un sitio Web de la Agencia en cuestión. Es recomendable que estos índices (*Harvest Control Lists*) estén escritos en XML, mejor que HTML, porque contienen campos específicos de información (al menos, título y URL de cada página clave) con una sintaxis semejante a esta:

```
<item>
<title> A Guide to the Australian Constitution </title>
<link> http://www.gov.au/constitution.htm </link>
</item>
```

³² Where and How Should AGLS Metadata be Stored and Accessed [documento HTML]. En: *Applying the AGLS Metadata Standard*. Canberra: Commonwealth of Australia, rev. 4 de Julio de 2000. Disponible en: http://www.naa.gov.au/recordkeeping/gov_online/agls/guidelines/chpt11.html#11.2 (consultado el 10 de agosto de 2000).

³³ *Metabrowser: A Web Browser for Metadata...*, Op. cit., http://metabrowser.spirit.net.au/manual/_Toc493945502.htm, y http://metabrowser.spirit.net.au/manual/_Toc493945536.htm

³⁴ Cfr. 4.3.2.2.

Con la utilización de estas HCL, los propietarios de los sitios Web pueden promocionar informaciones básicas tales como las páginas de búsqueda y las páginas que aglutinan recursos en forma de directorio.

Cfr. Harvest-Harvesting /Gather-Gathering.

Information brokering

Vid. Brokering, Broker.

Information gateway

Vid. Subject gateway/Information gateway.

Interoperability

*La interoperabilidad en informática se define, en general, como la habilidad que tiene un sistema o producto para trabajar con otros sistemas o productos sin un esfuerzo especial por parte del cliente*³⁵. Este concepto tiene una importancia creciente a tenor de las tecnologías de la información en el sentido de que la frase "*la red es el ordenador*" se está convirtiendo en realidad.

A pesar de la complejidad de este concepto y de sus múltiples implicaciones para los sistemas de recuperación de información basados en metadatos, lo hemos explicado en el apartado 8.3.2., al hablar de la necesidad de integración de metadatos, tanto a nivel técnico como a nivel semántico, para conformar bibliotecas híbridas o para facilitar un acceso global entre bibliotecas digitales. Lo recogemos también en este glosario para justificar su traducción por "interoperabilidad" en vez de interoperatividad. Aunque lo hemos encontrado en algunos casos³⁶ traducido por "interoperatividad", su traducción más directa

³⁵ Esta es la definición que aparece en el Diccionario especializado en informática What?is: http://searchsystemsmanagement.techtarget.com/sDefinition/0,,sid20_gci212372,00.html Citada también por Eric Miller. Interoperability: What is it and Why Should I Want it? [documento HTML] *Ariadne*, 26-Jun-2000, issue 24. Disponible en: <http://www.ariadne.ac.uk/issue24/interoperability/intro.html> (consultado el 8 de agosto de 2000).

³⁶ V. gr. Cristina Faba Pérez, Félix de Moya Anegón. Bibliotecas digitales: concepto y principales proyectos. *Investigación Bibliotecológica*, enero/junio 1999, vol. 13, nº 26, p. 66.

en español partiendo de que la palabra operatividad está aceptada en el DRAE como *capacidad de realizar una función*³⁷ nos ha parecido más preciso mantener la forma más parecida en inglés, una vez más, a través de la creación terminológica por calco semántico (en este caso, añadiéndole el prefijo *inter-*).

Vid. Crosswalks/Crossroads, Mapping.

Mapping

A pesar de que es un término muy utilizado en el contexto de la información automatizada, es difícil encontrar una definición exhaustiva de esta palabra en los glosarios y vocabularios constituidos para explicar la terminología informática. Según el vocabulario especializado para los estudios de bibliodocumentación³⁸ este término se puede traducir, en el contexto de la gestión de la información, como *clasificación, distribución u organización*. En el glosario que consigna Alschuler³⁹ en su obra sobre SGML, se define el concepto *mapping* —desde nuestro punto de vista y en el contexto de la materia que estamos tratando de forma muy acertada— como *una correspondencia programada entre información en diferentes formatos, usada para la traducción y la conversión de datos*. Ésta es la acepción que utilizamos en este trabajo, tomando la adaptación al español "mapeo" como término para referirnos a la conversión entre distintos formatos de metadatos.

Vid. Crosswalks/Crossroads, Interoperability.

Mark-up/markup language

Los lenguajes de marcado, o simplemente el marcado de documentos, es un concepto fundamental dentro de esta tesis doctoral. Por ello resulta indispensable, además de las precisiones que hemos consignado en el capítulo 5, establecer en este glosario su significado de forma inequívoca. Suele traducirse como lenguaje de marcas o también marcado ó balizado según se

³⁷ *Diccionario de la lengua española. Op. cit.,* <http://buscon.rae.es/drae/drae.htm>

³⁸ Antonio Lozano Palacios. *Op. cit.*, p. 217.

³⁹ Liora Alschuler. *Op. cit.*, p. 397.

haga el calco del inglés (*mark*) o del francés (*balise*). Lozano⁴⁰ lo traduce como etiquetado de documentos, Peis y Moya⁴¹ como lenguaje de etiquetado o marcaje, y Abaitua⁴² como maquetación, en el sentido de que originalmente, *markup* se utilizaba para designar el proceso de marcar un documento manuscrito con instrucciones acerca de cómo se deben utilizar los tipos de letra, los tamaños, los espacios, el sangrado etc.

Tim Berners-Lee habla de toda una familia de lenguajes de marcado, y define el *markup*, de una forma muy elocuente, como *descripción de documentos*⁴³, tal y como reflejamos en el capítulo 5. Todas estas expresiones traducidas, e incluso el término en inglés, son las que se han utilizado para referirnos aquí a los *markup languages*.

Metadata [Format/Model/Registry/Schema/Standard]

Como hemos explicado en otras partes de esta investigación, a lo largo de este trabajo intercambiamos con cierta espontaneidad las denominaciones formato, modelo, esquema y norma o estándar de metadatos para evocar un conjunto de elementos que conforman una especificación (normalmente asociada a un dominio informativo) de las etiquetas que se deben utilizar con un valor "deíctico" sobre un DLO. Ahora es el momento de proponer una definición unívoca para todos ellos, a la par que de subrayar algunos matices.

⁴⁰ Antonio Lozano Palacios. *Op. cit.*, p. 218. Sin embargo, en la versión electrónica de este diccionario <<http://eubd1.ugr.es/ris/risweb.isa>> incluye dentro de la voz *markup* las expresiones *markup code* que traduce como *etiqueta*, *símbolo*, *marca*, y *markup system*, como *sistema de marcas*. *Método para la identificación de las partes de un documento (título, parte, sección, párrafo, etc.) según una estructura estándar para su posterior proceso automatizado*.

⁴¹ Eduardo Peis, Felix de Moya. *Op. cit.*, p. 4-16.

⁴² Joseba Abaitua. La maquetación. En: *Material de referencia para un curso de introducción a SGML* [documento HTML]. Bilbao: Universidad de Deusto, rev. 8 de mayo de 1997. Disponible en: <http://sirio.deusto.es/abaitua/konzeptu/sgml/etiqueta.htm> (consultado el 3 de marzo de 2001).

⁴³ Tim Berners-Lee. *Tejiendo... Op. cit.*, p. 39.

Polydoratou y Nicholas⁴⁴ analizan las denominaciones más populares para designar los modelos de metadatos a través de un cuestionario distribuido, entre el 12 y el 15 de noviembre de 2000, en el segundo taller de SCHEMAS y en diferentes listas de discusión, de distintos ámbitos de implementación de metadatos (sector industrial, editorial, educativo, universitario, investigación y otros). Los resultados reflejaron que los términos más utilizados para referirse a los formatos de metadatos son: *schemas* y *standards* (20%), *element sets* (19%) y *formats* (16%), aunque también se utilizan otras denominaciones como *systems* (12%), *catalogues* (11%) y otros (2%).

Estos mismos autores⁴⁵ distinguen entre *metadata format*, como conjunto de reglas que gobiernan la estructura y contenido de los datos, y *metadata schema*, un conjunto estructurado de atributos con nombres de los elementos y una semántica asociada. Por el contrario, la definición que hace la Iniciativa de Metadatos del Dublin Core (DCMI) de *scheme/schema*⁴⁶ vuelve a justificar la asimilación de ambas denominaciones (*format* y *schema*): *combinación sistemática y ordenada de elementos. Un conjunto de reglas, mantenidas por una comunidad específica de usuarios, para codificar información.*

El DCMI también otorga este sentido a la expresión *metadata registry*, que podría dar lugar a equívocos en la traducción a través de lo que se denominan "falsos amigos" del inglés, y traducirlo como registro de metadatos (*Metadata record*, en realidad —*Vid. infr.*—).

⁴⁴ Panayiota Polydoratou, David Nicholas. Familiarity With and Use of Metadata Formats and Metadata Registries amongst Those Working in Diverse Professional Communities Within the Information Sector. *ASLIB Proceedings*, 2001, vol. 53, n° 8, p. 317, Figure 6.

Para argumentar la utilización de uno u otro término, podríamos haber realizado un muestreo genérico en la Web a través de la información reflejada en los motores de búsqueda, semejante al que realizamos para argumentar el uso de los términos "digital library" "electronic library" y "virtual library" (Graf. 7, 8, 9 y 10); pero nos ha parecido más oportuno citar los datos obtenidos por Polydoratou y Nicholas ya que son datos actualizados (2001) y que responden a una metodología más precisa (cuestionario directo realizado a profesionales) que la búsqueda indiscriminada por palabra clave que nos permiten los buscadores.

⁴⁵ *Ibid.*, p. 310.

⁴⁶ Mary S. Woodley. Glossary. *Op. cit.* <http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml#S> *Cfr. infr. Scheme.* La asimilación de los dos términos (*scheme/schema*) al mismo concepto, proviene de la DCMI, aunque también encontramos la confusión de ambos términos en otros autores. A pesar de que la traducción es la misma en ambos casos (esquema), es necesario hacer una diferenciación entre ambos términos que matizamos en las entradas *Schema/Scheme* de este mismo glosario.

Metadata registry debe traducirse como "modelo de metadatos", tal como hemos hecho en este trabajo, aludiendo a *un sistema accesible públicamente que almacena la semántica, estructura y formato de intercambio de un tipo de metadatos*⁴⁷. Sin embargo, la denominación de *metadata registry* proviene del estándar internacional ISO 11179 (*Specification and Standardization of Data Element*) diseñado para proporcionar un marco robusto para la definición de elementos de datos de forma persistente y sin ambigüedades entre las distintas comunidades de usuarios. La ISO 11179 es una amplia norma constituida por seis partes que empezó a desarrollarse en 1998 y comporta muchos más aspectos de la estructuración y el intercambio de información que los metadatos. En el marco de los metadatos, un registro recoge las definiciones oficiales de los términos que componen un estándar de metadatos, e incluso puede incluir políticas sobre el uso de los términos definidos.

Otra justificación posible para esta denominación "*registry*" (oficina de registro, literalmente) podría verse en el hecho de que la evolución, mantenimiento y desarrollo de un modelo de metadatos se encarga a una autoridad formal o una agencia que se responsabiliza de las políticas relativas al contenido y funcionamiento de un determinado modelo; *v. gr.* el mantenimiento y desarrollo del Dublin Core se encarga la DCMI, el CSDGM se encarga al *Federal Geographic Data Commitee* que da también nombre a dicho esquema de metainformación (FGDC).

Como ocurre en toda disciplina o tema que está gestándose científicamente, como es el caso de los metadatos y las bibliotecas digitales, la terminología no es estable, lo que justifica que se utilicen (y que utilicemos) distintos términos para evocar un modelo de metadatos. Por ejemplo se han empleado también, para referir a la realidad que estamos caracterizando aquí, los términos *template* o *plantilla*⁴⁸, e incluso *protocolo*⁴⁹). No obstante, si tuviésemos que aproximar

⁴⁷ *Ibid.*, <http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml#M>

⁴⁸ María Pinto, por ejemplo habla de *plantillas o formatos de metadatos semiestructurados* para referirse a los esquemas de metadatos, como por ejemplo el Dublin Core. María Pinto Molina. Tratamiento de los contenidos en la Sociedad de la Información. En: *La Sociedad de la Información: política...* *Op. cit.*, p. 278.

⁴⁹ Algunos autores como Vidal y Salvador se refieren a estos modelos de estructuración de la información como *protocolos*. Francisco Javier Vidal Bordés y José Antonio Salvador Oliván. *Op. cit.*, p. 198. Desde nuestro punto de vista, referirse a los esquemas de metadatos como *protocolos* no es correcto. Aunque partiendo de una definición genérica de protocolo como el conjunto especial de reglas que sirven para la comunicación entre sistemas en una conexión telemática, podríamos pensar [cont.]

todas estas denominaciones, en rigor, sólo deberían intercambiarse modelo y registro por un lado, y formato y esquema por otro, pudiendo referirnos a todos ellos (registro, modelo, formato y esquema) como estándares de metadatos en el sentido que hemos discutido en el apartado 6.2. de esta tesis doctoral.

Cfr. Scheme, Template.

Metadata record

Siempre que hemos utilizado este concepto en nuestro trabajo, nos hemos referido a "registro de metadatos", en el mismo sentido que se entiende "registro" en la información bibliográfica convencional. Más concretamente, siguiendo el glosario de la DCMI⁵⁰, podemos definir un registro de metadatos como *una representación, correcta desde el punto de vista sintáctico, de la información descriptiva (metadatos) para un recurso de información.*

Si un registro bibliográfico es la representación codificada de la información descriptiva de un documento, sea ésta en MARC o en formato ISBD, un registro de metadatos es la representación codificada de la información de un DLO o recurso electrónico, independientemente de que ésta se encuentre en HTML o XML, almacenada en la propia cabecera del documento o de forma externa.

Cfr. Metadata [Format/Model/Registry/Schema/Standard].

Metadata Registry

Vid. Metadata [Format/Model/Registry/Schema/Standard].

que un modelo de metadatos define también un conjunto de reglas que, al ser utilizadas en un contexto de Red, también facilitan el intercambio de información a través de la estructuración de la misma. Sin embargo, un protocolo es una descripción formal de las reglas que dos ordenadores deben seguir para intercambiar información, y la utilización de este término debe ser exclusiva de la interconexión entre máquinas, no de la definición, legible por el hombre además de por la máquina, que implica un esquema o formato de metainformación.

⁵⁰ Mary S. Woodley. Glossary. *Op. cit.*, <http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml#M>

Metatag

La comprensión de este término no presenta ningún inconveniente ya que los dos componentes que lo forman, el prefijo *meta-* (meta) y la palabra *tag* (etiqueta), tienen un significado perfectamente coherente en español: metaetiqueta. A lo largo de toda la tesis hemos utilizado, en algunas ocasiones, el término metaetiqueta como sinónimo de metadatos para simplificar su significado sobre todo en lo relativo a la creación de éstos (*Vid.* Capítulo 4). No obstante, como ya hemos advertido⁵¹, la problemática en la utilización de esta palabra, desde nuestro punto de vista, se debe a que la correspondencia semántica entre metadato y metaetiqueta no es total.

Una diferenciación específica entre ambos términos la encontramos en los conceptos de la página Web del software para gestión de metadatos Metamanage. Según se entiende en este software, los *metadatos son información clasificatoria. En relación con los sitios Web y con las Intranets, los metadatos se refieren a la información almacenada dentro del fichero HTML que describe información clave sobre el contenido del archivo. Normalmente se definen en un estándar denominado 'schema', que determina cómo deben usarse los metadatos*⁵²; mientras que las metaetiquetas *son unidades de información que residen dentro de la cabecera de un documento y que describen algo sobre dicho documento*⁵³.

Así pues, como advertíamos en el capítulo 4, todos los metadatos de una página HTML son metaetiquetas, mientras que no todas las metaetiquetas son metadatos (por ejemplo la etiqueta <META> con el atributo `http-equiv`, aunque es una información sobre la información ya que establece una comunicación entre la página Web y el servidor http, no es una información sobre el contenido de la página). Podemos hacer además una precisión desde el punto de vista formal: una metaetiqueta es el elemento de marcado (HTML o XML) que se utiliza para señalar e identificar los metadatos relativos a un DLO, mientras que metadatos pueden ser también clasificaciones temáticas u ontologías desde el punto de vista del contenido de un documento.

⁵¹ *Cfr.* Capítulo 4, nota 97.

⁵² *MetaManagement Concepts* [documento HTML]. Metamanagement, rev. 26 de julio de 1999. Disponible en: <http://www.metamanage.com/doco/concepts.htm#metatag> (consultado el 14 de abril de 2001). *Cfr.* 4.3.2.2. Igualmente se diferencia entre *schema/scheme*, *Vid. infr.*

⁵³ *Ibid.*

Model

Vid. Metadata [Format/Model/Registry/Schema/Standard].

Namespace

La expresión *namespace* es propia de XML y, por ello, también de RDF. Su traducción más adecuada sería *espacio de nombre*⁵⁴, aunque a lo largo de este trabajo utilizamos esa denominación, alternándola con su equivalente en inglés.

En la programación en C++ también existe este concepto, y en Java se denomina *package* a los espacios de nombre. Un *namespace es un dispositivo de alcance, utilizado para identificar unívocamente entidades registradas. Se pueden distinguir entidades llamadas igual en diferentes namespaces. Un namespace se identifica a través de un URI que incluye una denominación (nombre) y una versión del identificador*⁵⁵.

Debe tener el formato URL: comienza por el protocolo <http://>, seguido de su nombre de dominio y etiquetas de *namespace* opcionales con el formato de un nombre, una descripción muy corta del espacio de nombres y un número de versión opcional (el URL no tiene que apuntar expresamente a ningún archivo real). Según esto, así son las declaraciones de *namespace* para la Sintaxis RDF y para el modelo de metadatos del DC:

```
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:dc="http://purl.org/dc/elements/1.1/#"
```

La utilización de espacios de nombre permite la definición de un elemento a través de su identificación unívoca con un URI; aunque una etiqueta, por ejemplo `title`, pueda aparecer en distintos conjuntos de metadatos. En términos generales, se puede pensar en cualquier conjunto cerrado de nombres como un *namespace*. Tal es el caso de un vocabulario controlado como el LCSH, un conjunto de metadatos, como el DC o un conjunto de todos los URLs en un dominio determinado; todos ellos pueden entenderse como *namespaces* que se gestionan por la autoridad asignada de esos conjuntos de términos.

⁵⁴ Así lo encontramos traducido, por ejemplo, en: Elizabeth Castro. *Guía de aprendizaje XML*. Madrid, etc.: Prentice Hall, 2001, p. 95 y ss.

⁵⁵ Glossary (Manjula Patel). *Op. cit.*, <http://www.schemas-forum.org/related/glossary.html#n>

Nesting

A este término se hace referencia en español utilizando, en algunas ocasiones, el barbarismo "nesteado" que adapta a nuestra lengua el participio presente en inglés *nesting*, para denominar la forma en la que los subelementos pueden contenerse dentro de elementos más amplios, que resultan en múltiples niveles de metadatos. En el ámbito de la metainformación, tiene un significado metafórico, pues *to nest* significa en inglés "anidar", por ello se utiliza para designar esa forma de anidamiento entre distintos niveles de metadatos.

Pathfinders

El significado original de *pathfinder* en inglés es el de explorador o pionero. Sin embargo el sentido que toma en esta investigación es el de una especie de sistema de control bibliográfico por materias. Según la definición que da CORC, proyecto en el que surge este término, *son páginas Web que consolidan enlaces para seleccionar recursos sobre una materia; normalmente categorizan los recursos y proporcionan anotaciones informativas*⁵⁶. Los *pathfinders* permiten a los usuarios acceder a recursos previamente seleccionados e incluyen también descripciones de materiales sobre ese tema que, aunque no estén accesibles en Internet, sí lo están en la biblioteca. Son algo así como directorios temáticos seleccionados o "de calidad". Según la terminología que utilizamos en el capítulo 7, se entienden como sistemas de recuperación de información de cariz bibliotecario que tratan de presentar a los usuarios los mejores recursos sobre un tema. Aunque es una denominación específica en sistema CORC, también se denominan *pathfinders*, con un significado semejante, a las guías creadas *ad hoc* por la *Internet Public Library* (IPL) para ayudar a los usuarios a iniciar una búsqueda sobre un campo temático particular, tanto en Internet como en una biblioteca local tradicional⁵⁷. En cualquier caso, los *pathfinders* se entienden como ayudas para la selección o la búsqueda temática de recursos.

Cfr. Subject gateway/Information gateway.

⁵⁶ OCLC CORC Systems Guides: Create and Use Pathfinders [documento HTML]. OCLC, rev. 1 de mayo de 2001. Disponible en: <http://www.oclc.org/corc/documentation/pathfinders.pdf> (consultado el 8 de mayo de 2001).

⁵⁷ Pathfinders (The Internet Public Library): <http://www.ipl.org/ref/QUE/PF>

Quotinformation

Esta palabra está formada a través de un *portmanteau*, es decir, una figura retórica que se presenta muy comúnmente en inglés como recurso de creación lingüística por composición cuando dos términos y sus significados se unen en una sola palabra nueva, cuyo sentido es una combinación de sus dos generadoras.

En este caso, las palabras que conforman este nuevo término son *quote* (citar, cita) e *information* (información) y su significado más sencillo sería el de "información citada" o "citar la información". Es un término utilizado por Iannella y Waugh⁵⁸ para definir los metadatos. Se nos ocurre argumentar que, de igual forma que en inglés se utiliza *quote X* para decir "citando las palabras de X", *quotinformation* puede caracterizar a los metadatos en el sentido de que éstos citan, resumen o aluden a la información contenida en un recurso electrónico.

Registry

Vid. Metadata [Format/Model/Registry/Schema/Standard].

Cfr. Metadata record.

Schema

Es un término de origen griego⁵⁹ [σχῆμα] cuyo uso en el ámbito de la metainformación se funda en la analogía con el mundo de las bases de datos, donde se utiliza para describir la estructura de una tabla en una base de datos relacional. No obstante, en sentido estricto, el término *schema* se utiliza en XML y RDF para definir vocabularios:

- En XML, un *schema* es un modelo para describir la estructura de la información de toda una clase de documentos. Los esquemas XML son

⁵⁸ Renato Iannella, and Andrew Waugh. *Metadata: Enabling...* Op. cit. <http://www.dstc.edu.au/RDU/reports/CAUSE97>

⁵⁹ Por eso en algunos casos encontramos el cultismo para el plural *schemata* (esquemas). Por ejemplo en: Ora Lassila. Web Metadata: A Matter of Semantics. *IEEE Internet Computing*, July-August 1998, p. 34.

patrones para estructuras y vocabularios XML y suponen una forma alternativa a las DTDs para definir dichas estructuras.

- En RDF, un *schema* define un mecanismo para declarar las propiedades de los elementos así como las relaciones entre las propiedades y los recursos, y también la semántica para hacer todo esto.

En sentido más genérico, aplicado a todo el ámbito de la metainformación, un esquema —sobre todo si está acompañado del determinante "de metadatos"— es el conjunto de elementos y las reglas que constituyen un modelo de metainformación. Como explica Hjelm⁶⁰ *un schema puede interpretarse como un acuerdo en un vocabulario común para una aplicación particular que implica el intercambio de documentos.*

La traducción del término en este trabajo, salvo cuando se refiere expresamente a esquemas XML o esquemas RDF, suele ir acompañada del término en inglés entre paréntesis: "esquema (*schema*)" por contraposición a "esquema (*scheme*)", cuando el contexto en que se presenta puede inducir a error o cuando es necesaria una diferenciación entre ambos términos.

Cfr. Scheme.

Vid. Metadata [Format/Model/Registry/Schema/Standard].

Scheme

Scheme es, en sentido estricto, un atributo propio de la etiqueta <META> del HTML que sirve para definir el esquema que se debe utilizar para interpretar el valor de una propiedad. El hecho de ser un atributo de HTML justificaría su uso en inglés, sin embargo, entraña un significado más complejo que un mero componente estructural de dicho lenguaje de marcado.

A pesar de que la traducción es la misma, "esquema", y que la DCMI⁶¹, identifica *schema* y *scheme*, es importante diferenciar la cobertura semántica de cada uno de estos términos en inglés, sobre todo por la diferenciación de uso

⁶⁰ Johan Hjelm. *Op. cit.*, p. 125.

⁶¹ Mary S. Woodley. Glossary. *Op. cit.* <http://dublincore.org/documents/2001/04/12/usageguide/glossary.shtml#S>

que hacen algunas aplicaciones informáticas. Tanto la página Web del software Metamanager de la empresa australiana Edime <www.edime.com.au>, como el glosario que incluye el manual del software, también australiano, Metabrowser⁶², se expresa una diferencia entre ambos términos. El concepto de *schema*, responde a lo que hemos tratado al hablar de modelos de metadatos, asimilable a un formato de metainformación. El concepto de *scheme* hace alusión un *fichero de tesauro controlado, una lista de valores posibles que puede tener una metaetiqueta en concreto*⁶³.

Según esta matización, un *schema* está formado, como anticipamos anteriormente, por los elementos y las reglas que constituyen un modelo de metadatos, y un *scheme* es el conjunto más amplio de valores que pueden tener esos elementos. Por ello, en algún caso hemos traducido *scheme* como estándar o esquema de contenido, por contraposición a *schema* como estándar o esquema de estructura semántica. Por ejemplo, en el proyecto Scorpion⁶⁴, se refieren a la Clasificación Decimal de Dewey como un *scheme*. Así pues, el DC es un *schema* de metadatos y para completar la información del elemento DC.Subject de este *schema* se puede utilizar el *scheme* DDC; la ISO 8601 para el formato de fecha es un *scheme*, mientras que el GILS, por poner otro ejemplo, es un *schema*. No obstante, a efectos de su traducción y utilización en este trabajo, hablaremos en ambos casos de "esquemas", completando su significado a través de sendos determinantes: esquemas "de metadatos", en el primer caso, y esquemas "de clasificación" o "de indización", en el segundo.

Aunque en esta tesis hemos adoptado la diferenciación entre *schema* y *scheme* que acabamos de señalar, y que parece ser la más consensuada en la bibliografía, encontramos expresiones confusas como la de Iaonid y

⁶² *Metabrowser: A Web Browser for Metadata. Version 1.1 for Microsoft Windows* [documento HTML]. Erindale: Metabrowser System, Spirit Consulting, 2001. Disponible en: <http://metabrowser.spirit.net.au/manual> (consultado el 14 de abril de 2001).

- Schema: http://metabrowser.spirit.net.au/manual/_Toc493945540.htm
- Scheme: http://metabrowser.spirit.net.au/manual/_Toc493945541.htm

⁶³ *MetaManagement Concepts. Op. cit.*, <http://www.metamanage.com/doco/concepts.htm#schema>

⁶⁴ Scorpion el proyecto de investigación de OCLC que tiene como objetivo la creación de herramientas para el reconocimiento automático de materias, basándose en esquemas de clasificación bien conocidos como la DDC. Más información, *Vid. The Scorpion Project* <<http://orc.rsch.oclc.org:6109>>; Keith Shafer. *Scorpion... Op. cit.*, p. 28-29.

Bowman⁶⁵, que hablan de *schemes* refiriéndose al DC y a las AACR y de *schema* de vocabulario controlado⁶⁶, o la asimilación de ambos conceptos, que ya hemos mencionado, en la *Dublin Core Metadata Initiative*.

Cfr. Schema.

Seamless (Web/searching)

La palabra en inglés *seamless* significa "sin costuras", o a lo sumo, buscando un significado en el contexto de la Documentación se puede traducir como uniforme, homogéneo, totalmente integrado, perfecto, sin notar la diferencia⁶⁷. En el ámbito de la recuperación de información en la Web, es recurrente el uso de este adjetivo para referirse a sistemas "perfectamente integrados"; por lo tanto éste es el sentido que le hemos dado tanto en la redacción del trabajo como en la aprehensión del concepto en la bibliografía consultada.

Además, la capacidad semántica de este adjetivo en inglés, en el ámbito de los servicios de información Web, se ha llevado incluso a denominar *Seamless* al informe realizado en la ciudad inglesa de Essex sobre la modernización de la administración identificando, la necesidad de desarrollar estándares y de fomentar la cooperación para mejorar el intercambio de información entre organizaciones y utilizar la tecnología para hacer más fácil a los ciudadanos y a las empresas comunicarse con la Administración, recibir información y servicios. Dicho informe dio lugar a un sistema de información, también denominado *Seamless* <<http://www.seamless.ac.uk>> y a un esquema (*scheme*) de metadatos propio formado por treinta y tres atributos así como a un tesoro basado en el utilizado en las bibliotecas de Essex y otras listas de autoridad como los nombres de lugares, etc. Con todo, *SEAMLESS*, como sistema de información original de dicha ciudad inglesa, responde al paradigma de biblioteca digital que se defiende en esta tesis⁶⁸, donde los metadatos son un

⁶⁵ Aurora Iaconid, Bibiana Bowman. Data and Metadata: An Overview of Organization in Searchable Full-Text Databases. En: *Creating Web-Accessible Databases...* *Op. cit.*, p. 144.

⁶⁶ *Ibid.*, p. 151.

⁶⁷ Antonio Lozano Palacios. *Op. cit.*, p. 308; <http://eubd1.ugr.es/RIS/risweb.isa>

⁶⁸ Para más información sobre el sistema SEAMLESS, Vid. Mary Rowlett, et al. *SEAMLESS: An Organisational and Technical Model for Seamless Access to Distributed Citizen's Information*. London: Library and Information Commission, 2000.

elemento normativo más del entramado político y filosófico de la democratización del acceso a la información a través de la tecnología.

Search buckets

El significado literal de esta expresión sería algo así como "cubo de búsqueda" pero, aplicado a la recuperación de información basada en metadatos, es un concepto que surge —aunque no es exclusivo de ella, como explicaremos— en la ADL⁶⁹, una biblioteca digital sobre recursos geoespaciales que trata de integrar distintos modelos de metadatos proporcionándoles parámetros comunes para la búsqueda.

Los *search buckets*, en el contexto de la *Alexandria Digital Library*, son *pequeños conjuntos de atributos de metadatos de alto nivel que pueden utilizarse para hacer que distintas colecciones puedan tener los mismos metadatos de búsqueda. Se designan explícitamente para la interrogación. Los campos de metadatos de un registro que subyace no se mapean uno a uno. Más bien, los grupos de campos podrían mapearse⁷⁰ a un bucket y a los campos se les puede hacer corresponder más de un bucket⁷¹*. Esos atributos de alto nivel son: *geographic location, type, format, topical text, assigned terms, originator, date range, identifier*. Constituyen un mecanismo propio de la ADL para crear los metadatos contenedores de diversos recursos y pretenden ser el núcleo de las búsquedas. En realidad, los *search buckets* podrían entenderse como un modelo *ad hoc* de metadatos de la *Alexandria Digital Library*, puesto que definen una serie de

⁶⁹ El concepto de *search buckets* se explica sobre todo en: James Frew, et al. Generic Query Metadata for Geospatial Digital Libraries [documento HTML]. En: *Conference IEEE Computer Society*, 1999. Disponible en: <http://computer.org/proceedings/meta/1999/papers/55/jfrew.htm> (consultado el 4 de junio de 2001); pero también aparece en otras publicaciones relativas a la ADL, v. gr.: Linda L. Hill, et al. Collection Metadata Solutions... *Op. cit.*, p. 1174-1175.

⁷⁰ *Cfr. supr., mapping.*

⁷¹ Katy Ginger. *DLESE Search Buckets: Details Search Buckets Concepts and Schemes for Interoperability* [documento HTML]. DLESE.org, rev. 4 de diciembre de 2000. Disponible en: <http://www.dlese.org/Metadata/buckets.htm> (consultado el 4 de junio de 2001). DLSE, utiliza el modelo de metadatos IMS, destinados para la educación; sin embargo, al ser una biblioteca digital para el sistema de educación de ciencias de la tierra, sus planteamientos en materia de metadatos están muy vinculados a la *Alexandria Digital Library*, de ahí que utilicen también el concepto de *search buckets*.

atributos clave. A diferencia de un esquema de metadatos convencional, los *search buckets* no sirven para homogeneizar las descripciones de una colección digital, sino para interoperar entre diversos recursos con metadatos distintos en una biblioteca digital heterogénea.

A pesar de que en este trabajo nos referimos al sentido que da la ADL a los *search buckets*, existen otras visiones, próximas aunque diferentes, en torno al concepto de *buckets*. La más interesante es la de Michael Nelson, un ingeniero de la NASA que desarrolló en su tesis doctoral⁷² toda una teoría sobre los *buckets* como *objetos inteligentes para las bibliotecas digitales*. Esta misma idea de su tesis, donde defiende los *buckets* para separar el contenido de la información de su almacenamiento y recuperación en aras a incrementar la interoperabilidad entre bibliotecas digitales, se recoge en un artículo posterior⁷³, donde se definen los *buckets* como *constructos acumulativos, inteligentes y orientados a objetos que contienen datos, metadatos y los métodos para acceder a ambos*.

Search engine

Los *search engines* se denominan en español, haciendo una traducción directa del inglés, como motores de búsqueda. Un motor de búsqueda, según Martijn Koster⁷⁴, es *un programa que busca entre distintos conjuntos de datos*. En el contexto de la Web, según reconoce Koster, la expresión "*search engine*" se utiliza para los formularios que permiten buscar en las distintas bases de datos de documentos acopiados por un software robot. En general, se denomina "motor de búsqueda" a toda la infraestructura de un tipo de servicio de recuperación de información en Internet que utiliza ese software para buscar recursos HTML que previamente un robot recopiló en una o varias bases de datos.

⁷² Michael L. Nelson. *Smart Objects for Digital Libraries. Ph.D Dissertation*. Old Dominion University, August 2000 [documento RTF]. Disponible en: http://mln.larc.nasa.gov/~mln/phd/?method=display&pkg_name=dissertation.pkg&element_name=daily/phd.doc (consultado el 14 de diciembre de 2000).

⁷³ Michael L. Nelson and Kurt Maly. Buckets: Smart Objects for Digital Libraries. *Communications on the ACM*, May 2001, vol. 44, n° 5, p. 60-62.

Para denominar estos sistemas de recuperación de información en Internet, existe una amplia terminología que tiene diferentes matices y que se funda en diversas metáforas, así, los motores de búsqueda se vienen llamando: arañas (*spiders*), orugas (*crawlers*), gusanos (*worms*), vagabundos (*wanders*), hormigas Web (*WebAnts*), etc., que con sus particularidades y diferencias, tratan de designar un sistema de búsqueda de información en la WWW. En muchos casos se llama a estos sistemas *robots* o *search engines*, haciendo una sinécdoque del concepto y denominando al todo (el sistema de recuperación de información compuesto por un robot que recorre la Web, una base de datos, un sistema de indización automática, un motor de búsqueda y una interfaz de consulta), por la parte (el *robot*, en unos casos o el *search engine*, en otros).

En esta tesis doctoral utilizamos el concepto de motor de búsqueda e incluso el propio término en inglés *search engine* para referirnos tanto en sentido estricto, al software de recuperación de una base de datos, como en el sentido convencional que se da a esta expresión, esto es, a toda la tipología de servicios de búsqueda que funcionan de la forma que hemos descrito.

Por último, la utilización, en algunos casos, del anglicismo (*search engine*) en lugar de su traducción exacta en español, persigue simplemente evitar la redundancia de términos en la exposición pues, como hemos dicho, la correspondencia en español es perfectamente traducible e interpretable en el contexto de este trabajo. Normalmente hemos tratado de usar la expresión en inglés para referirnos expresamente al software de recuperación de una base de datos, más que al servicio genérico de búsqueda de la Web, que denominamos predominantemente con la expresión en español "motor de búsqueda".

Vid. Searchbot.

SearchBot

Dentro de las múltiples denominaciones que reciben los *search engines*, los términos *Knowbot* y *Searchbot* —basados en la composición *know* (conocimiento) / *search* (búsqueda) + *bot* (abreviatura de robot)— aluden generalmente, a sistemas de indización y recuperación de información que trabajan como agentes cliente que, una vez instalados en el computador del

⁷⁴ Martijn Koster. The Web Robots FAQ... [documento HTML] The Web Robots Pages, rev. 10 de julio de 2001. Disponible en: <http://www.robotstxt.org/wc/faq.html#engine> (consultado el 7 de noviembre de 2001).

usuario, realizan búsquedas automáticas con una autoridad delegada de sus usuarios, a cuyos resultados se añaden otras prestaciones como la organización, la posibilidad de acceso *off-line*, etc. A este tipo de agentes de búsqueda pertenecen aplicaciones como Copernic <<http://www.copernic.com>> o Hurricane Web <<http://www.gatecomm.com>>, que son además, metabuscadores.

La traducción que hacemos de ambos términos, no es directa como en el caso de *search engine*, sino conceptual: agentes de búsqueda, por contraposición a los robots que funcionan en modo servidor y que sólo permiten la búsqueda por palabras clave a través de la interrogación directa.

Vid. Search engine.

Serendipity

Se trata de un concepto utilizado esporádicamente en la bibliografía sobre recuperación de información. Desde nuestro punto de vista es muy preciso para denominar aquellos hallazgos fortuitos de información, ya sea recorriendo secuencialmente las estanterías de una biblioteca o librería, ya sea siguiendo enlaces en Internet. *Serendipity* es, *un proceso creativo en los sistemas de recuperación de información y en particular en las bibliotecas digitales*⁷⁵ que supone una forma poco ortodoxa pero a veces muy efectiva de recuperación de información.

La utilización del término *serendipity* se atribuye al novelista inglés Horace Walpole (1717-1797), quien acuñó el término para designar: *la facultad de hacer, accidentalmente, descubrimientos oportunos e inesperados*⁷⁶. En este trabajo utilizamos el término "serendipidad", transliterando a la forma nominal en español ante la falta de una traducción conveniente, en el contexto de la técnica del *browsing* o navegación, refiriéndonos al tipo de recuperación de información en Internet que se realiza por azar a través del seguimiento de

⁷⁵ Eleaine G. Toms. *Op. cit.*, http://www.ercim.org/publication/ws-proceedings/DelNoe01/3_Toms.pdf

⁷⁶ *Serendipity Homepage* [documento HTML]. Ohio: SouthEastern Ohio Regional Freenet Ohio University-Governor's Scholars, 18 de julio de 1996, rev. 26 de julio de 1996. Disponible en: <http://www.seorf.ohiou.edu/~xx045/hide/serendip/serendip.htm> (consultado el 18 de julio de 1999).

enlaces o *links*, pero se puede ampliar la utilización de este término a todo el contexto de la recuperación de información en el espacio digital.

Spam, Spamming



En un diccionario de inglés convencional, v. gr. el Collins⁷⁷, el único significado que tiene la voz *Spam* es una marca americana registrada de carne de cerdo, con lo cual, su significado traído aquí no tendría ningún sentido ni aplicación. Incluso en los nuevos glosarios de informática surgidos a raíz de la Red, así como en diccionarios actualizados (como la herramienta lexicográfica contextual Babylon Translator), se dan distintos significados al término *spamming*, vinculados generalmente al envío masivo de correos electrónicos con fines comerciales⁷⁸. En el contexto de esta investigación utilizamos este término — como en otros casos, por ausencia de traducción en nuestra lengua, de la forma anglosajona— en el ámbito de la recuperación genérica de información en la WWW. Lee Anne Phillips define *spamming* como *la práctica por medio de la cual los vendedores sin escrúpulos de la Web rellenan sus páginas con palabras que no tienen relación, con la intención de recuperarlas en respuesta a cualquier búsqueda*⁷⁹.

Se trata de una técnica que consiste básicamente en repetir, dentro de una etiqueta <META> una palabra clave, de tal forma que el registro del documento en cuestión suba al inicio de una lista de resultados de un buscador de Internet.

⁷⁷ Colin Smith. *Op. cit.*, p. 1478.

⁷⁸ Concretamente, la aplicación *Babylon Translator*, define el sustantivo en inglés *spamming*, como: *envío de avisos en el Internet sobre un tema poco importante a un número grande de grupos de discusión (en el mismo momento)*, y *spam* como correo electrónico "basura". Babylon: <http://babylon.com> (desde este mismo URL se puede descargar la aplicación cliente de este diccionario-traductor contextual; fichero: babylon32.exe, 18..Kb).

⁷⁹ *Web spamming —a practice whereby unscrupulous web hucksters lard their pages with unrelated words in a scheme to bring up their page in response to any search at all—*. Lee Anne Phillips. *Using XML: Special Edition*. Bestseller ed. Indianapolis: Que, 2000, p. 490. Si analizamos la terminología en inglés que utiliza Phillips para definirlo podemos encontrar una justificación a la exótica denominación de este concepto: *Lard* en inglés, significa mechar o rellenar con tocino; quizás la marca de carne *Spam* sea una marca específica de carne para rellenar y por eso se utiliza este término para designar el "relleno" de información con fines promocionales en los resultados de los buscadores genéricos.

Esta práctica se ha convertido en habitual debido a que muchos de los motores de búsqueda realizan un cálculo de relevancia, y por ende, ordenan los resultados, según el número de veces que una palabra se utiliza en una página, dándole mayor importancia cuando se repite en determinadas partes del documento, por ejemplo, en la cabecera o <HEAD> donde se colocan los metadatos. La práctica del *keyword-spamming* puede consistir también en introducir palabras que son irrelevantes para un sitio determinado con el fin de atraer visitantes bajo falsos pretextos. Dentro de las técnicas habituales de *spam/spamming* aplicadas a la edición Web, podemos destacar:

- Repetición o empleo excesivo de palabras clave, normalmente en la etiqueta <META NAME="keywords">.
- Utilización de *keywords* que no reflejan el contenido del sitio.
- Utilización de la técnica del *refresh*⁸⁰ rápido.
- Utilización de texto en el mismo color que el background, de tal forma que el usuario no puede ver la repetición de descriptores pero los motores de búsqueda sí indizan dicha información textual.
- Duplicación de páginas con URLs diferentes.
- Utilizar diferentes páginas que enlazan o dirigen al mismo URL.

El *spamming* es la alteración o la creación de un documento con la intención de engañar a un sistema de indización y almacenamiento en Internet. Algunos creadores de páginas Web utilizan algunas de estas técnicas subversivas para conseguir que sus sitios aparezcan más altos en la lista de resultados. Sistemas como Infoseek, por ejemplo penalizan esta práctica excluyendo de su índice las páginas que intentan subir su visibilidad a través del *spamming*. Incluso, repeticiones extremas de *spam*, en este sentido pueden ser causa de acciones legales.

⁸⁰ Esto implica la utilización de la metaetiqueta de comunicación con el sistema <META HTTP-EQUIV="refresh" CONTENT="1"; url="nombredelfichero.htm"> que permite reenviar en un segundo de la página principal (v. gr. index.html) a otra (nombredelfichero.htm) de tal forma que la primera contiene información "falsa" para mejorar el ranking de relevancia de los buscadores.

Stakeholder

Aunque es muy común el uso de este término en la literatura científica en inglés, no sólo en el ámbito de la Biblioteconomía, la Informática o los metadatos, su significado habitual es "tenedor de apuestas", "accionista"⁸¹ o "parte implicada, participante, protagonista"⁸². Sin embargo, en cualquier discusión de origen científico, hablar de *stakeholders* implica una persona o grupos de personas con un interés o pretensión común, que tienen la posibilidad de generar influencia, crear opinión o marcar la dirección de un proyecto o iniciativa.

En el ámbito de la metainformación, y así lo hemos traducido y utilizado, son aquellos agentes (personas o instituciones) implicadas en el desarrollo de un modelo de metadatos cuya opinión y liderazgo es decisivo en la adopción y evolución de un estándar/es. Por ejemplo, desde un punto de vista individual, Stu Weibel o Eric Miller son *stakeholders* del DC, o Dan Brickley y Ralph Swick de RDF; a nivel institucional, son *stakeholders*, por ejemplo, la Library of Congress, en el desarrollo del EAD, o de forma colectiva también, se puede considerar *stakeholder* a la lista *www-rdf-interest*, promovida en el seno del W3C que participa activamente en el desarrollo del estándar de metadatos RDF.

Standard

Vid. Metadata [Format/Model/Registry/Schema/Standard].

Subject gateway/Information gateway

El término *gateway* (*vid. supr.*), cuyo significado es "pasarela", unido a la palabra información, podría traducirse como "pasarela de información" o "pasarela de materia/temática". Sin embargo su traducción da lugar a una expresión sincategoremática en español, esto es, una frase cuyo significado no es la yuxtaposición del significado de sus componentes. Otras denominaciones, también sincategoremáticas, utilizadas en inglés para designar la realidad del

⁸¹ Antonio Lozano Palacios. *Op. cit.*, p. 328

⁸² Antonio Lozano Palacios. *Op. cit.*, <http://eubd1.ugr.es/RIS/risweb.isa>

servicio de información que evocan son⁸³: *subject-based information gateways (SBIGs)*, *subject-based gateways*, *subject index gateways*, *virtual libraries*, *clearinghouses*, *subject trees* o *pathfinders*, etc. o incluso *quality controlled subject services*⁸⁴.

Una *subject gateway* es, según Campbell⁸⁵:

Un mecanismo basado en la Web para acceder a una colección de alta calidad, formada por recursos que pueden ser útiles para la investigación en una disciplina particular, donde los recursos son evaluados y descritos por especialistas de la información, tales como bibliotecarios científicos.

En el sentido que se le da en esta tesis doctoral una *information/subject gateway* es un servicio de información temática donde los recursos están seleccionados según un criterio de calidad de la información, y organizados siguiendo esquemas de análisis documental (indización, resumen, clasificación y/o metadatos). Básicamente, una *subject gateway* es un tipo de biblioteca digital/virtual distribuida, formada por un conjunto o colección de recursos de información electrónica que se ajustan a una serie de criterios —por ejemplo, un área temática, una institución creadora o criterios de tipo geográfico y/o lingüístico— que han sido previamente seleccionados y organizados siguiendo algún esquema de metadatos. Esa circunstancia de "selección y organización" es lo que nos hace considerar a este tipo de servicios de recuperación de información en Internet, como un tipo especial de biblioteca digital, o para muchos, biblioteca virtual⁸⁶.

⁸³ John Kirriemuir, et al. Cross-Searching Subject Gateways: The Query Routing and Forward Knowledge Approach [documento HTML]. *D-Lib Magazine*, January 1998, vol. 4, nº 1. Disponible en: <http://www.dlib.org/dlib/january98/01kirriemuir.html> (consultado el 21 de septiembre de 1999).

⁸⁴ Traugott Koch. *Browsing and Searching...* *Op. cit.*, http://www.ub2.lu.se/nav_menu.html

⁸⁵ Debbie Campbell. *Op. cit.*, <http://www.vala.org.au/vala2000/2000pdf/Campbell.PDF>

⁸⁶ *Cfr.* 8.1. para la diferenciación de estos términos. Como explicamos en el capítulo 8, algunos autores utilizan la expresión *biblioteca virtual* para referirse a aquella que no posee la colección a la que da acceso. Para nosotros, como matizamos al hablar de la Biblioteca Virtual *Miguel de Cervantes*, el calificativo de "virtual" responde por un lado a la tecnología que utiliza para constituirse y por otro a su alcance de carácter universal, en la propia abstracción del concepto. *Cfr.* 9.2.1., Fig. 62. No obstante reconocemos que una buena expresión de este concepto en español es la de: *bibliotecas virtuales temáticas*, que podemos fundamentar en la denominación (*bibliotecas virtuais temáticas*) que

[cont.]

Renardus⁸⁷, el proyecto de colaboración entre pasarelas temáticas europeas, distingue entre *information* y *subject gateway*, diciendo que *information gateway* es el término genérico que refiere a toda la variedad de guías Web dirigidas a recursos de información de Internet; mientras que *subject gateway* son guías para la recuperación de recursos basados en una materia que proporcionan enlaces a documentos, colecciones, sitios o servicios, normalmente accesibles en Internet. La característica principal de estas guías es la clasificación por materias y la descripción de los recursos. Según esta diferenciación, podemos decir, por ejemplo, que Infomine es una *information gateway* e incluso también ODP; mientras OMNI y SOSIG son *subject gateways* y bibliotecas digitales distribuidas según nuestro concepto que funda la determinación de "biblioteca digital" en la organización y descripción de recursos a través de metadatos.

Aunque a lo largo de esta investigación utilizamos mayoritariamente la expresión en inglés *subject gateway*, o biblioteca virtual temática. También utilizamos su traducción directa "pasarela de materia" reconociendo la poca precisión que tiene este término en español, pero relegando otras traducciones que se han hecho de él, como portal o agencia de evaluación⁸⁸ que nos parecen demasiado amplias o menos oportunas para designar la realidad de las colecciones distribuidas y organizadas en Internet que tratamos de reflejar en esta investigación.

Vid. Gateway.

Cfr. Clearinghouse, pathfinders.

Surrogate, Subrogate

Al igual que ocurre con otros términos, *surrogate* (s.) y *subrogate* (v.) no tienen una correspondencia semántica en español, a pesar de que sí exista traducción para ellos: sustituto o reemplazante y subrogar, respectivamente.

adoptó el CNPq brasileño para designar estos sistemas de información digital. *Vid.* <http://www.prossiga.br/bvtematicas>

⁸⁷ *Gateways Defined* (Renardus): <http://www.renardus.org/gateway/gateway.html> *Cfr.* Capítulo 8, nota 145.

⁸⁸ *Cfr.* 8.2.2.

Subrogar es un término que se utiliza en español, en el ámbito jurídico con el significado de sustituir. Sin embargo, en el ámbito de la Documentación hablar de "sustitutos" es demasiado vago y no corresponde a ningún concepto bibliotecario identificado. La utilidad que tiene un sustituto (*surrogate*) es determinar el grado en el que puede localizarse un documento (por autor, título, materia); identificar y describir el objeto de información e identificar, asimismo, su ubicación concreta. Burnett, Bor Ng y Park⁸⁹ dicen, además, que un sustituto o subrogado es la unidad básica de un sistema de catalogación de una biblioteca, que representa el objeto de información. Con estas prerrogativas asociadas al vocablo y concepto inglés *surrogate* podemos determinar que los metadatos son sustitutos o "subrogados" de un documento entendido como objeto (DLO), y así lo hemos expresado en este trabajo.

Syndication/ syndication format

En general, *syndication* significa sindicato, con lo cual, esta expresión *syndication format* se podría traducir como "formato de sindicato". Como en otros casos reflejados en este glosario, el significado convencional de este término no tiene ninguna relación semántica con el ámbito de la recuperación de información en la Web. Sin embargo, en Internet, ha cobrado un nuevo sentido cuando se utiliza al lado de la palabra formato. *Syndication format* implica la distribución o redistribución de contenido en otros sitios Web. Los formularios típicos de contenido incluyen titulares, noticias, imágenes y clips multimedia (vídeo y/o audio). Los sitios Web que producen contenido lo colocan en otros para generar que se reconozca la marca o para provocar un mayor tráfico en el sitio. Si bien el concepto está claro, la traducción es complicada, por ello hemos utilizado agregación o afiliación de contenido e incluso sindicación de contenidos, adoptando el valor que tiene el término en francés para referirse a la publicación de una misma noticia en varios periódicos⁹⁰.

Existe además una aplicación XML específica para definir el nivel de agregación o sindicación llamada OCS (*Open Content Syndication*). El formato

⁸⁹ Kathleen Burnett, Kwong Bor Ng, Soyeon Park. *Op. cit.*, p. 1210.

⁹⁰ Para aspectos prácticos sobre la sindicación de contenidos, *Vid.* Thomas Pack. *Go Forth and Syndicate: Five Questions to Ask When Getting Started*. *Econtent*, August 2001, vol. 24, n° 6, p. 24-30.

de directorio OCS se ha diseñado para permitir la construcción de listados, legibles por máquina, de canales que se usarán en los portales, en los software cliente que sirven para el intercambio de titulares y en otras aplicaciones similares⁹¹.

Template

La traducción de *template* es plantilla, en el sentido de documento utilizado como diseño básico para nuevos documentos. El uso del término *template* con este significado parte de las aplicaciones de Microsoft, donde se pueden utilizar distintas plantillas para construir documentos semejantes (cartas, faxes, *curricula*, etc.). En algunas ocasiones también se denominan hojas de estilo⁹², entendidas como un fichero o formulario utilizado en los procesadores de texto y en autoedición, donde se definen las características de presentación de un documento.

En el ámbito de los metadatos, las plantillas hacen alusión a los formularios Web que completará el usuario con los datos relativos a la publicación electrónica, según distintos campos que se ajustan a un modelo o esquema de metadatos, y por ello en algunos casos se utiliza *template* como una metonimia del propio formato de metadatos. No obstante, el significado que hemos dado en nuestra tesis al término *template/plantilla* es el de herramienta informática preconcebida en forma de formulario que facilita la creación o asignación de metadatos *online* según un modelo determinado.

La complejidad de estas plantillas puede ir desde la facilidad de crear las etiquetas más simples que contemplan algunos motores de búsqueda para su indización —tipo `<META NAME="keywords">`, `<META NAME="description">` (Figs. 4, 6-9, 11)—, a plantillas que siguen formatos y esquemas de metainformación de propósito general o específico, más complejos (Figs. 23, 27, 28 y 30), tal y como se ha explicado en el apartado 4.3. Si bien es cierto que los software cliente para la creación de metainformación frecuentemente adoptan también este aspecto de plantillas, hemos utilizado el término *template* o plantilla para designar aquellas herramientas para la creación de metadatos que basan su funcionamiento en un formulario Web.

⁹¹ Más información sobre OCS: <http://internetalchemy.org/ocs/directory.html>

⁹² *Webopedia... Op. cit.*, http://webopedia.internet.com/TERM/s/style_sheet.html

Vid. Metadata [Format/Model/Registry/Schema/Standard].

Web, WWW

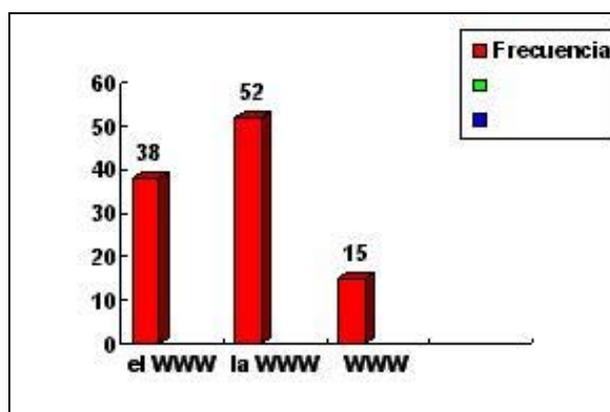
La aparición del término *Web* en un glosario de estas características puede parecer elemental, ya que su uso y significado están ampliamente difundidos, no entrañan ninguna dificultad y están universalmente aceptados. Incluso, su nivel de aceptación en nuestro idioma ha hecho que en la última edición del DRAE (presentada en el *II Congreso Internacional de la Lengua Española*, el 17 de octubre de 2001), se haya introducido el término *web*, aunque con un significado muy vago (*red informática*⁹³), como término aceptado en español.

Su traducción general como “telaraña” o “telaraña mundial” (WWW), ha quedado relegada a expresiones *cuasi*-literarias y la aceptación del anglicismo es internacional tanto para evocar servidores como páginas Web o, en general, para hacer alusión a esta nueva forma de información hipertextual. El hecho de recoger “Web” en este glosario tiene que ver más con la inserción morfológica y sintáctica de este término en nuestra lengua que con su significado y/o traducción, que no conllevan ninguna duda o imprecisión e incluso, como hemos dicho, están formalmente aceptados en ella.

En cuanto al género de este anglicismo, consideramos más correcto el femenino en el sentido de que telaraña en español es femenino, y en el caso de que se utilice con un sentido que evoque “página Web” heredará igualmente el género del sustantivo (página). Así se utiliza mayoritariamente en la propia Web (Graf. 11) y ese es el género que le ha asignado la Real Academia Española⁹⁴.

⁹³ La entrada concreta del DRAE sobre la voz Web es exactamente ésta: (Del ingl. *web*, red, malla). 1. **f.** *Inform.* Red informática. Además se ha incluido también *página web*, definida como: *documento situado en una red informática, al que se accede mediante enlaces de hipertexto*. *Diccionario de la lengua española*. *Op. cit.*, <http://buscon.rae.es/drae/drae.htm>

⁹⁴ *Ibid.*



Graf. 11. Formas de aparición del anglicismo WWW⁹⁵

No obstante, cabe la posibilidad de que Web sobrentienda la expresión "servidor, sitio o sistema Web", en cuyo caso podría atribuírsele el género masculino. Hay quienes opinan que no tiene género; pero su uso extendido hace que se utilice tanto "el/un Web", como "la/una Web" y por ello, en esta tesis pueden aparecer ambas formas en cuanto al género⁹⁶, pero damos preeminencia al femenino —la Web/WWW— por ser el correcto, según el DRAE, en español.

En relación con el número del vocablo WWW/Web, hemos utilizado siempre el singular aplicando el plural, en su caso, en el sustantivo que acompañe (p. ej. "páginas Web").

Finalmente, en cuanto a su utilización en minúsculas (www/web) o en mayúsculas (WWW/Web), hemos preferido estas últimas. A pesar de que hay muchos detractores del abuso de las letras capitales que se introducen por

⁹⁵ Fuente: Pedro José Sampedro Losada. *Anglicismos, barbarismos, neologismos y "falsos amigos" en el lenguaje informático* [documento HTML]. ATI, Grupo de Lengua e Informática, rev. 6 de junio de 2000. Disponible en: <http://www.ati.es/gt/lengua-informatica/externos/sampedr1.html#grafi1> (consultado el 2 de agosto de 2001).

⁹⁶ En la versión española del libro de Tim Berners-Lee with Mark Fischetti. *Weaving the Web...* en español, *Tejiendo...* *Op. cit.*, (Vid. Capítulo 2, nota 43) al que hemos hecho alusión en varias ocasiones a lo largo de este trabajo, su traductora, Mónica Rubio Fernández, transcribe siempre "el" Web, por ello, cuando citamos literalmente esta obra utilizamos el género masculino. Sólo hemos reconocido la alternativa de utilizar "Web" con el artículo masculino, en la expresión "*promoción de la Web*" o "*promoción de la Web*" en el sentido del carácter genérico y no vinculante que puede evocar la expresión en masculino *Cfr.* fundamentalmente, 4.3.1.

influjo del inglés en nuestra lengua⁹⁷, hemos elegido las mayúsculas, por una parte, para reconocer su importancia e identidad a través de un signo gráfico, y por otra, porque se trata de una sigla en inglés (World Wide Web).

Webby

Webby en el contexto de Internet es el nombre que recibe el premio de la Academia Internacional de Artes y Ciencias Digitales <<http://www.webbyawards.com>>, y es también un software para aumentar el tráfico de sitios Web <<http://www.webby.com>>. El hecho de que sea un nombre propio de alto contenido comercial⁹⁸, no sería razón para incluirlo en un glosario de estas características, pero sí lo es su uso como adjetivo. Como *quotinformation* (Vid. *supr.*), es también un *portmanteau* intraducible en español, cuyo significado responde a una combinación de la semántica implícita de las palabras que lo forman *Web* y *by* (cerca, al lado, por). Quiere decir algo así como "adecuado para la Web", y es el calificativo que utiliza Dumbill⁹⁹ para referirse a RDF en comparación con XML. Megginson¹⁰⁰, por

⁹⁷ El influjo del inglés en nuestra lengua está haciendo que ésta esté, cada vez más, inundada de mayúsculas innecesarias. El inglés tiene, *v. gr.* una regla ortográfica que obliga a escribir en mayúsculas todas las palabras de un título (así hacemos en la bibliografía de esta tesis ya que la norma ISO 690-1987, establece —apartado 6.3.— que, *para el uso de las mayúsculas, se seguirá el uso ortográfico de la lengua en que se da la información*). Asimismo la lengua inglesa utiliza las mayúsculas en la primera palabra de siglas y acrónimos (como en el caso de CPU, *Central Processing Unit*, o de la propia WWW), que al traducirse en español, absorben la misma tendencia (Unidad Central de Proceso, en vez de unidad central de proceso) aunque no se trate de nombres propios, ni de instituciones. Muchos autores critican esta práctica en español, por ejemplo, según Amado de Miguel, *la decisión de poner con mayúsculas las palabras que describen cosas es un rasgo infantiloides que no ayuda mucho a la coherencia del idioma*. Amado de Miguel. *La perversión del lenguaje*. Madrid: Espasa Calpe 1985. Citado por Pedro José Sanpedro Losada. *Ibid.*, <http://www.ati.es/gt/lengua-informatica/externos/sampedr2.html#mayusculas>

⁹⁸ *Webby* es también por ejemplo, el nombre que recibe un software para la televisión de entretenimiento interactiva ("*intertainment*") en Internet. Vid. <http://www.webbycorp.nl>

⁹⁹ Edd Dumbill. *Op. cit.*, [http://weblogs.oreillynet.com/edd/discuss/msgReader\\$65](http://weblogs.oreillynet.com/edd/discuss/msgReader$65) Vid. 5.2.3.

¹⁰⁰ *The two best candidates for representing objects right now are RDF and XMI (since they're both controlled by independent bodies). Personally, I prefer RDF for data exchange, because it's more Webby (XMI is just UML with tags instead of little pictures)*. David Megginson. Re: Dissillusioned about Interoperability [mensaje de correo en lista de distribución]. *Xml-dev*, 8 de octubre de 1999; 08.11. Disponible en: <http://www.lists.ic.ac.uk/hypermail/xml-dev/xml-dev-Oct-1999/0101.html> (consultado el 7 de noviembre de 1999).

su parte, se refiere a RDF en estos mismo términos "*more webby*", al compararlo con XMI. El uso en inglés de *webby* se está generalizando, por ejemplo, cuando una empresa elige un acrónimo o un nombre que lo haga más proclive a formar parte de un dominio Web, ese nombre comercial se califica de *webby*¹⁰¹.

Workshop

Aunque la traducción de *workshop*, como "taller" es clara y no induce a error, a lo largo de esta tesis, hemos utilizado tanto el término en inglés, como su significado en español. La adopción y utilización del anglicismo se debe a que, si bien en el ámbito latinoamericano es bastante común denominar a estos seminarios de trabajo monográficos "talleres", el uso en nuestro país no es habitual y podría resultar un tanto anómalo si no intercambiásemos el término en inglés con el término en español.

Los *workshops*, como se ha podido comprobar en el texto de este trabajo y también en la bibliografía, son una forma habitual de encuentros presenciales de discusión en el ámbito de los metadatos, constituyendo —en algunos casos, como por ejemplo en el DC/DCMI— verdaderas series a través de las cuales se va desarrollando un estándar o modelo de metainformación.

WWW

Vid. Web.

¹⁰¹ *Vid.* Sarah Bedford. A Tangled WEB: An Industry Reinventing Itself, Internet Consulting Hunkers Down for the Long Haul [documento HTML]. *Business Forward*, May 2000, issue 25. Disponible en: <http://www.bizforward.com/wdc/archives/2001-05/buzzguide> (consultado el 7 de noviembre de 2001).

ÍNDICE DE SIGLAS Y ACRÓNIMOS

Sigla Significado/ URL

- 5PM** V Programa Marco de IDT
(Fifth Framework Programme)
<http://www.cordis.lu/fp5/home.html>
Vid. IDT, FP
- AACR2** Anglo-American Cataloging Rules
- AAP** Association of American Publishers
<http://www.publishers.org>
- AAT** Arts & Architecture Thesaurus
<http://www.getty.edu/research/tools/vocabulary/aat>
- ACH** Association for Computing in the Humanities
<http://www.ach.org>
Vid. TEI
- ACL** Association for Computational Linguistics
<http://www.cs.columbia.edu/~acl/home.html>
Vid. TEI
- ACM** Association for Computing Machinery
<http://www.acm.org>
- A-CORE** Metadata about content Metadata
<http://www.metadata.net/ac/draft-iannella-admin-01.txt>
- ADAM** Art, Design, Architecture & Media Information Gateway
<http://www.adam.ac.uk>
- ADEPT** Alexandria Digital Earth Prototype
<http://www.alexandria.ucsb.edu/adept>
- ADL** Alexandria Digital Library
<http://www.alexandria.ucsb.edu>
- ADT** Australian Digital Theses Project
<http://adt.caul.edu.au>
- AGIFT** Australian Governments' Interactiva Functional Thesaurus
http://www.naa.gov.au/recordkeeping/gov_online/agift/summary.html
- AGLS** Australian Government Locator Service
http://www.naa.gov.au/recordkeeping/gov_online/agls/summary.html
- AGNIC** Agriculture Network Information Center
<http://www.agnic.org>
- AGRIGATE** Agriculture Information Gateway for Australian Researchers
<http://www.agrigate.edu.au>
- AHDS** Arts and Humanities Data Service
<http://ahds.ac.uk>
- ALA** American Library Association
<http://www.ala.org>

Sigla Significado/ URL

- ALCTS** Association for Library Collections & Technical Services
<http://www.ala.org/alcts>
Vid. ALA
- ALIWEB** Archie-Like Indexing for the Web
<http://aliweb.emnet.co.uk>
<http://www.aliweb.com>
- ALLC** Association for Literary and Linguistic Computing
<http://www.kcl.ac.uk/humanities/cch/allc>
Vid. TEI
- ANSI** American National Standard Institute
<http://www.ansi.org>
- ANZLIC** Australia New Zealand Land Information Council
<http://www.anzlic.org.au>
<http://www.anzlic.org.au/asdi/metaelem.htm>
- APA** American Psychological Association
<http://www.apa.org>
- API** Application Programming Interface
- ARL** Association of Research Libraries
<http://www.arl.org>
Vid. DID
- ARP** Another RDF Parser
<http://www-uk.hpl.hp.com/people/jjc/arp>
- ASI** American Society of Indexers
<http://www.asindexing.org>
- ASN.1** Abstract Syntax Notation One (ISO 8824)
<http://asn-1.com>
Vid. XER
- ATI** Asociación de Técnicos de Informática
<http://www.ati.es>
- AU** Agente de Usuario (navegador, *browser*)
- AVEL** Australian Virtual Engineering Library
<http://avel.edu.au>
- AVELSD** Australian Virtual Engineering Library Sustainable Development
http://avel.library.uq.edu.au/avel_sd
- BATS** BEP AGLS Tagging System (Ver. 0.9, Oct.2000)
bep.support@dewrsb.gov.au
- BEP** Business Entry Point
<http://www.business.gov.au>
<http://www.ausindustry.gov.au/documents/dir20/doc507420.pdf>

Sigla Significado/ URL

- BIB-1** (Conjunto de atributos utilizado en la formulación de una pregunta en un cliente Z39.50 para especificar la búsqueda)
Nueva formulación de atributos BIB-1
<http://www.loc.gov/z3950/agency/defns/bib1.html>
Vid. Z39.50-2001
- BIBCO** Bibliographic Record Component
<http://www.loc.gov/catdir/pcc/bibco.html>
Vid. PCC
- BibHit** Catalogue of Internet Resources Indexed by Using Metadata (Proyecto de danés)
<http://www.bibhit.dk/info/english.htm>
- BIBLINK** (Proyecto europeo en el marco de Telematics for Libraries)
<http://hosted.ukoln.ac.uk/biblink>
- BIS** Business Information System
- Biz/ed** Business Education on the Internet
<http://www.bized.ac.uk>
- BNE** Biblioteca Nacional de España
<http://www.bne.es>
- BSWL** Basic Semantic Web Language
<http://purl.org/net/bswl>
<http://infomesh.net/2001/07/bswl>
- BSZ** Bibliotheksservice-Zentrum Baden-Württemberg
<http://www.bsz-bw.de>
- BUBL** The Bulletin Board for Libraries
<http://bubl.ac.uk>
- BVC** Biblioteca Virtual *Miguel de Cervantes*
<http://www.cervantesvirtual.com>
- CARMEN** Content Analysis, Retrieval and MetaData: Effective Networking
<http://www.mathematik.uni-osnabrueck.de/projects/carmen>
- CaSIL** California Spatial Information Library
<http://www.gis.ca.gov>
Vid. CERES
- CATRIONA** Cataloguing and Retrieval of Information Over Networks Applications
<http://bubl.ac.uk/org/catriona/cat1rep.htm>
CATRIONA II
<http://wp269.lib.strath.ac.uk:5050/Cat2/index.html>
- CBU** Control Bibliográfico Universal
- CC/PP** Composite Capability/Preference Profiles: Structure and Vocabularies
<http://www.w3.org/TR/CCPP-vocab>

Sigla Significado/ URL

- CC:DA** Committee on Cataloging: Description and Access
<http://www.ala.org/alcts/organization/ccs/ccda/ccda.html>
Vid. ALA
- CCS** Computing Classification System (ACM)
<http://www.acm.org/class/1998/TOP.html>
Vid. ACM
- CDF** Channel Definition Format/Framework
<http://www.w3.org/TR/NOTE-CDFsubmit.html>
- CDLI** Cuneiform Digital Library Initiative
<http://cdli.ucla.edu>
Vid. DLI2
- CDLR** Centre for Digital Library Research
<http://cdlr.strath.ac.uk>
- CDWA** Categories for the Description Works of Art
<http://www.getty.edu/research/institute/standards/cdwa>
- CEDARS** CURL Exemplars in Digital Archives
<http://www.leeds.ac.uk/cedars/index.htm>
Vid. CURL
- CELT** Center for Electronic Text in the Law
<http://www.law.uc.edu/CETL>
- CEN** European Committee for Standardization
Comité Européen de Normalisation
<http://www.cenorm.be>
- CEN/ISSS** CEN/ Information Society Standardization System
<http://www.cenorm.be/iss>
- CENDI** (Grupo de trabajo de científicos, técnicos y gestores de información de las Agencias Federales Americanas)
http://www.dtic.mil/cendi/cendi_home.html
http://www.dtic.mil/cendi/proj_biblio_metadata_stds.html
http://www.dtic.mil/cendi/proj_cont_vocab.html
- CERES** California Environmental Resources Evaluation System
<http://ceres.ca.gov>
California Metadata Management System
<http://ceres.ca.gov/catalog/metadata.html>
Vid. CaSIL
- CERIF** Common European Research Information Format
<http://www.cordis.lu/cerif>
- CETH** Center for Electronic Text in the Humanities
<http://www.ceth.rutgers.edu>
- CGI** Common Gateway Interface
- CIA** Consejo Internacional de Archivos
Vid. ICA

Sigla Significado/ URL

- CICA** Centro Informático Científico de Andalucía
<http://www.cica.es>
- CIDOC** International Committee for Documentation of the International Council of Museums
Comité International pour la Documentation du Conseil International des Musées
<http://www.cidoc.icom.org>
Vid. ICOM
- CIMI** Consortium for the Computer Interchange of Museums Information
<http://www.cimi.org>
Estándares CIMI: <http://www.cimi.org/publications.html>
- CNI** Coalition for Networked Information
<http://www.cni.org>
- CIP** Cataloging In Publication
Vid. ECIP
Common Indexing Protocol
<http://www.faqs.org/rfcs/rfc2651.html>
Vid. RFC 2651
- CIRCA** Conceptual Information Retrieval and Communication Architecture (Tecnología propietaria de Applied Semantics, Inc.)
http://www.appliedsemantics.com/as_solutions_tech.shtml
- CLIR** Council on Library and Information Resources
<http://www.clir.org>
- CNPq** Conselho Nacional de Desenvolvimento Científico e Tecnológico
<http://www.cnpq.br>
- CoBRA+** Computerized Bibliographic Record Actions
<http://portico.bl.uk/gabriel/cobra>
- CoI** Community of Interest
(concepto emanado de los productos de tSA, Klarity y BetaMeta):
http://www.klarity.com.au/whitepaper/meta-data_white_paper-05.htm#P64_6215
<http://www.topic.com.au/products/bettameta.html>
Vid. CVU
- CONSER** Cooperative Online Serials (*Vid. PCC*)
<http://lcweb.loc.gov/acq/conser>
- COPAC** CURL (*Vid.*) Online Public Access Catalogue
<http://copac.ac.uk/copac>
- CORBA** Common Object Request Broker Architecture
<http://www.corba.org>
- CORC** Cooperative Online Resource Catalog Project (OCLC)
<http://www.oclc.org/corc>
<http://corc.oclc.org>

Sigla Significado/ URL

- COVAX** Contemporary Culture Virtual Archive in XML
<http://www.covax.org>
- CR** Candidate Recommendation (W3C)
- CRM** CIDOC object-oriented Reference Model (CRM)
http://cidoc.ics.forth.gr/docs/crm_version_3.0.rtf
Vid. CIDOC
- CSCMM** Content Standard for Computer Model Metadata
<http://www.geog.ucsb.edu/~scott/metadata/standard/standard.doc>
- CSDGM** Content Standard for Digital Geospatial Metadata
<http://www.fgdc.gov/metadata/contstan.html>
<http://www.fgdc.gov/metadata/csdgm>
Vid. FGDC
- CSIRO** Commonwealth Scientific & Industrial Research Organization
<http://www.csiro.au>
- CSS** Cascading Style Sheets
<http://www.w3.org/Style/CSS>
- Cultivate** European Cultural Heritage Network (Europa e Israel)
<http://www.cultivate-eu.org>
- CURL** Consortium of University Research Libraries
<http://www.curl.ac.uk>
- CVU** Comunidad Virtual de Usuarios (Redes temáticas)
<http://www.rediris.es/cvu>
Vid. CoI
- CWA** CEN Workshop Agreement
Vid. CEN
- CYRUS** Classify Your Resources Using Scout
Vid. ISP
- DAMA** Data Management Association
<http://www.dama.org>
- DAML** DARPA Agent Markup Language
<http://www.daml.org>
- DARPA** Defense Advanced Research Projects Agency
<http://www.darpa.mil>
- DBMS** Data Base Management Systems
Vid. SGBD
- DC** Dublin Core
<http://purl.org/DC>
Desde el 16/01/01 reemplazado por DCMI: <http://dublincore.org>

Sigla Significado/ URL

DC-AP	Dublin Core – Application Profiles
DCD	Document Content Description for XML http://www.w3.org/TR/NOTE-dcd
DC-Lib	Dublin Core Library Application Profile http://www.dublincore.org/documents/library-application-profile
DCMES	Dublin Core Metadata Element Set http://dublincore.org/documents/1998/09/dces <i>Vid. Z39-85</i>
DCMI	Dublin Core Metadata Initiative http://dublincore.org http://uk.dublincore.org (mirror) http://au.dublincore.org (mirror)
DCq	Dublin Core Qualifiers/Qualified http://dublincore.org/documents/2000/07/11/dcmes-qualifiers
DCRDF	Dublin Core Resource Description Framework http://purl.org/DC http://dublincore.org
DDC	Dewey Decimal Classification http://www.oclc.org/dewey
DDI	Data Documentation Initiative http://www.icpsr.umich.edu/DDI <i>Vid. ICPSR</i>
DDL	Description Definition Language <i>Vid. MPEG7</i>
DEC	Data Element Concept (ISO 11179)
Delos	Networks on Excellence on Digital Libraries http://www.ercim.org/delos
DESIRE	Development of a European Service for Information on Research and Education http://www.desire.org
DGXIII	Dirección General XIII de la Comisión Europea Information Society Directorate-General http://europa.eu.int/comm/dgs/information_society/index_en.htm
DID	ARL Digital Initiatives Database http://www.arl.org/did <i>Vid. ARL</i>
DIF	Directory Interchange Format
DIG	Digital Imaging Group http://www.digitalimaging.org

Sigla Significado/ URL

- DIG35** Digital Imaging Group Metadata Initiative
http://www.digitalimaging.org/i_dig35.html
- DIO** Document-like Information Object
- DIS** Draft International Standard
Vid. ISO
- DK5** Sistema de clasificación Danés
<http://www.dbb.dk/kataloger/dk5.asp>
- DLESE** Digital Library for Earth System Education
<http://www.dlese.org>
- DLF** Digital Library Federation
<http://www.clir.org/diglib>
Vid. CLIR
- DLI** Digital Library Initiative
<http://www.dli2.nsf.gov/dlione> (Phase 1: 1994-1998)
<http://www.dli2.nsf.gov/index.html> (Phase 2: 1999-)
- DLO** Document Like-Object
- DNS** Domain Name System
- DOI** Digital Object Identifier
<http://www.doi.org>
Syntax for the Digital Object Identifier, *Vid.* Z39.84
- DOM** Document Object Model
<http://www.w3.org/DOM>
- DONOR** Directory Of Netherlands Online Resources
<http://www.kb.nl/coop/donor/index-en.html>
- DRAE** Diccionario de la Real Academia Española
<http://buscon.rae.es/drae/drae.htm> (22^a ed., 2001)
- DR-LINK** Document Retrieval using Linguistic Knowledge
<http://www.mnis.net>
- DSIG** Domain Special Interest Group (Statistics)
- DSTC** Distributed Systems Technology Centre
<http://www.dstc.edu.au>
- DTD** Document Type Definition/Declaration
Definición/Declaración del Tipo de Documento
Vid. SGML, XML
- EAD** Encoded Archival Description
<http://lcweb.loc.gov/ead>

Sigla Significado/ URL

- EARL** The Evaluation and Report Language
<http://www.w3.org/2001/03/earl>
- EBX** Electronic book exchange
<http://www.ebxwg.org>
- ECDL** European Conferences on Digital Libraries
ECDL 1997: <http://www.iei.pi.cnr.it/ErcimDL/dl3.html>
ECDL 1998: <http://www.ics.forth.gr/2EuroDL>
ECDL 1999:
<http://www.cse.unsw.edu.au/dblp/db/conf/ercimdl/ercimdl99.html>
ECDL 2000: <http://www.bn.pt/org/agenda/ecdl2000>
ECDL 2001: <https://www.ecdl2001.org>
ECDL 2002: <http://www.ecdl2002.org>
- ECIP** Electronic Cataloging In Publication
- EDI** Electronic Data Interchange
- EDMS** Electronic Document Management Systems
- EdNA** Educational Network Australia
<http://www.edna.edu.au/EdNA>
Estándar de Metadatos propio basado en el DC:
<http://standards.edna.edu.au/metadata>
- EELS** Engineering E-Library, Sweden
<http://eels.lub.lu.se>
- EEVL** Edinburgh Engineering Virtual Library
<http://www.eevl.ac.uk>
- e-GIF** Electronic Government Interoperability Framework
<http://www.govtalk.gov.uk/interoperability/interoperability.asp>
- EIO** Electronic Information Objects
Vid. DLO, DIO
- eLib** The Electronic Libraries Programme
<http://www.ukoln.ac.uk/services/elib>
eLib Web Page Metadata
<http://www.ukoln.ac.uk/services/elib/metadata.html>
Listado de proyectos de eLib
<http://www.ukoln.ac.uk/services/elib/projects>
- ELISE** Electronic Library Image Service for Europe
http://nile.dmu.ac.uk/elise/e2_intro.html
- EML** Ecological Metadata Language
<http://knb.ecoinformatics.org/software/eml>
- EOR** Extensible Open RDF
<http://eor.dublincore.org>

Sigla Significado/ URL

- EPA** Environmental Protection Agency
<http://www.epa.gov>
Vid. CENDI
- ER** Electronic Resource/s
- ERCIM** European Research Consortium on Informatics and Mathematics
<http://www.ercim.org>
- ERIC** Educational Resources Information Center
<http://www.eric.ed.gov>
- ETSI** European Telecommunications Standards Institute
<http://www.etsi.org>
- EUN** European School Net
- FAQ** Frequently Asked Questions
- FASTER** Flexible Access to Statistics, Tables and Electronic Resources
<http://www.faster-data.org>
- FDBS** Federated Data Base System
- FEDORA** Flexible and Extensible Digital Object and Repository Architecture
<http://www.cs.cornell.edu/cdlrg/FEDORA.html>
- FGDC** Federal Geographic Data Committee
<http://www.fgdc.gov>
Vid. CSDGM
- FIGIT** Follett Implementation Group in Information Technologies
http://www.jisc.ac.uk/pub/c4_94.html
- FIPA** Foundation for Intelligent Physical Agents
<http://www.fipa.org>
<http://www.fipa.org/specs/fipa00011/XC00011B.html>
- FOLDOC** Free Online Dictionary of Computing
<http://foldoc.doc.ic.ac.uk/foldoc/index.html>
- FP** Framework Program (UE)
FP4 (1994-1998): http://www.cordis.lu/es/src/f_002_es.htm
FP5 (1998-2002): <http://www.cordis.lu/fp5/home.html>
- FRBR** Functional Requirements for Bibliographic Records
<http://www.ifla.org/VII/s13/frbr/frbr1.htm>
- FTP** File Transfer Protocol
- G7** Grupo de los siete países más industrializados: Estados Unidos, Canadá, Francia, Alemania, Italia, Gran Bretaña y Japón

Sigla Significado/ URL

- G8** G7 + Rusia (desde 1998)
- GABRIEL** Gateway to Europe's National Libraries
<http://www.bl.uk/gabriel>
- GCMD** The Global Change Master Directory (NASA)
<http://gcmd.gsfc.nasa.gov>
- GDI+** GDI Plus (propuesta de metadatos para imagen de Microsoft)
http://www.microsoft.com/hwdev/display/GDIplus_Metadata.htm
- GEM** Gateway to Educational Materials
<http://www.geminfo.org>
<http://www.thegateway.org>
- GEMCat** GEM Catalogue
http://www.geminfo.org/Workbench/Workbench_cataloging.html
- GEMET** General Multilingual Environmental Thesaurus
http://www.mu.niedersachsen.de/cds/etc-cds_neu/software.html#GEMET
- GIF** Graphics Interchange Format
- GILS** Global Information Locator Service
<http://www.gils.net>
<http://www.gils.net/elements.html>
- GINF** Generic Interoperability Framework
<http://www-diglib.stanford.edu/diglib/ginf>
- GIS** Global Information Society
Geographical Information Systems
Vid. SIG
- GKP** Global Knowledge Partnership
<http://www.globalknowledge.org>
- GLO** Grail-Like Object
- GML** Generalized Markup Language
- GOL** Government On Line
- GPN** Global Knowledge Partnership
<http://www.globalknowledge.org>
- GRS** Generalized Record Syntax
- GTEL** Georgia Tech Electronic Library
<http://www.library.gatech.edu>

Sigla Significado/ URL

- HCL** Australian Government Harvest Control Lists-(Metabrowser)
http://www.naa.gov.au/recordkeeping/gov_online/agls/guidelines/chpt11.html#11.2
- HCU** Humanities Computing Unit
<http://www.hcu.ox.ac.uk>
- HEDS** Higher Education Digitisation Service
<http://heds.herts.ac.uk>
- HFN** Human Friendly Names
(El debate sobre esta forma de identificación de recursos Web se inició en el RFC 2276)
<http://www.ietf.org/rfc/rfc2276.txt?number=2276>)
- HILT** Hight Level Thesaurus
<http://hilt.cdlr.strath.ac.uk>
<http://www.ukoln.ac.uk/metadata/hilt>
- HISoftware** Hiawatha Island Software Company
<http://www.hisoftware.com>
- HPCC** High Performance Computing and Communications
<http://www.ccic.gov/pubs/imp95>
- HSTF** Hisoftware Subject Tree Format
(Extensión de los archivos que conforman las listas de materia que maneja la aplicación TagGen a través de su Hisoftware Subject Tree Manager).
Vid. HISoftware.
- HTML** HyperText Markup Language
<http://www.w3.org/MarkUp>
Vid. XHTML
- HTTP** HyperText Transfer Protocol
<http://www.w3.org/Protocols>
Vid. RFC 2616
- HyLife** Hybrid Libraries for the Future
<http://hylife.unn.ac.uk>
- IAF** Information access, filtering, analysis and handling
<http://www.cordis.lu/ist/ka3/iaf/index.htm>
Vid. IST
- IAFA** Internet Anonymous FTP
<http://aliweb.emnet.co.uk/iafa.txt>
- IBO** Information Bearing Object
- ICA** International Council of Archives
<http://www.ica.org>
- ICE** Information and Content Exchange Protocol
<http://www.w3.org/TR/NOTE-ice>

Sigla Significado/ URL

ICOM	International Council of Museums http://www.icom.org <i>Vid. CIDOC</i>
ICPSR	Inter-University Consortium for Political and Social Research http://www.icpsr.umich.edu/ <i>Vid. DDI</i>
ICRA	Internet Content Rating Association (Asociación de Clasificación de Contenidos de Internet) http://www.rsac.org
IDA	Integrated Document Access Projects
IDT	Investigación y Desarrollo Tecnológico
IE	Internet Explorer (Microsoft)
IEC	International Electrotechnical Commission http://www.iec.ch
IEEE	Institute of Electrical and Electronics Engineers http://www.ieee.org
IESG	The Internet Engineering Steering Group http://www.ietf.org/iesg.html <i>Vid. IETF</i>
IETF	Internet Engineering Task Force http://www.ietf.org
IFLA	International Federation of Library Associations and Institutions http://www.ifla.org http://ifla.inist.fr (<i>mirror</i> europeo)
IHR-info	Institute of Historical Research http://ihr.sas.ac.uk
IMesh	International Collaboration on Internet Subject Gateways (International Meshes) http://www.imesh.org
IMS	Instructional Management Systems http://www.imsproject.org
IMT	Internet Media Types (MIME Types) <i>Vid. MIME, y RFC 2376</i>
INDECS	INteroperability of Data in E-Commerce Systems http://www.indecs.org http://www.indecs.org/pdf/framework.pdf
Infoset	XML Information Set http://www.w3.org/TR/xml-infoset <i>Vid. XML</i>

Sigla	Significado/ URL
Inspec	Engineering and Information Science Scholarly Database http://www.iee.org.uk/publish/inspec
InterCat	Internet Cataloging Project (OCLC) http://www.oclc.org/oclc/man/catproj/catcall.htm
InterPARES	International Research on Permanent Authentic Records in Electronic Systems http://is.gseis.ucla.edu/us-interpares
IP	Internet Protocol
IPL	Internet Public Library http://www.ipl.org
IR	Information Retrieval <i>Vid.</i> RI
IRC	Internet Relay Chat
IRLT	Institute for Learning and Research Technology http://ilrt.org
IRP	Internationally Registered Profile
IRS	Information Retrieval System/s
IRTF	Internet Research Taskforce http://www.irtf.org
IRTF-RD	Internet Research Task Force-Resource Discovery Group http://www.irtf.org/charters/resource-discovery.html
Isaac	Isaac Network http://scout.cs.wisc.edu/research/isaac
ISAAR (CPF)	International Standard Archival Authority Record for Corporate Bodies, Persons and Families http://lettere.unipv.it/obc/add/infap/archdes/isaar_e.html
ISAD (G)	General International Standard Archival Description http://lettere.unipv.it/obc/add/infap/archdes/isad(g)e.html
ISBD	International Standard Book Description
ISBD (CF)	International Standard Book Description (Computer Formats)
ISBD (ER)	International Standard Book Description (Electronic Resources)

Sigla Significado/ URL

ISBN	International Standard Book Number http://www.isbn.org
Iscares	Information landscapes <i>Vid.</i> ADL
ISI	Institute for Scientific Information
ISO	International Standards Organization http://www.iso.ch
ISP	Internet Scout Project http://scout.cs.wisc.edu
IST	Information Society Technologies http://www.cordis.lu/ist IST in FP6: http://www.cordis.lu/ist/fp6/fp6consult.htm
IT	Information Technologies
JBO	JITI Builder Office (HiSoftware) http://www.hisoftware.com/Franklin/JITIBLD.htm <i>Vid.</i> JITI
JEP	Journal of Electronic Publishing http://www.press.umich.edu/jep
JISC	Joint Information System Committee http://www.jisc.ac.uk
JITI	Just In Time Information http://www.hisoftware.com/jiti.htm <i>Vid.</i> HiSoftware
JoDI	Journal of Digital Information http://jodi.ecs.soton.ac.uk
JSTOR	Journal STORage Project http://www.jstor.org/jstor <i>Vid.</i> UMDL, DLI
JPEG	Joint Photographic Expert Group <i>Vid.</i> JPG
JPG	Joint Photographic Expert Group (extensión de archivos gráficos) <i>Vid.</i> JPEG
LC	Library of Congress http://lcweb.loc.gov
LCC	Library of Congress Classification http://www.loc.gov/catdir/cpsolcco/lcco/lcco.html

Sigla Significado/ URL

- LCSH** Library of Congress Subject Headings
<http://lcweb.loc.gov/cds/lcsh.html>
- LDAP** Lightweight Directory Access Protocol
- LDR** Leader
Vid. MARC
- LEAD** Live Early Adoption and Demonstration
<http://www.w3.org/DesignIssues/Architecture.html#Collaboration>
Vid. W3C
- LETRS** Library Electronic Text Resource Service
<http://www.indiana.edu/~letrs>
<http://www.letrs.indiana.edu>
- LII** Librarians Index to Internet
<http://lii.org>
- LISA** Library & Information Science Abstracts
- LITA** Library and Information Technology Association
<http://www.lita.org>
Vid. ALA
- LL** Liaison Librarian
- LOM** Learning Object Metadata
<http://ltsc.ieee.org/doc/wg12/LOM3.8.html>
Vid. LTSC
- LTSC** Learning Technology Standards Committee
<http://ltsc.ieee.org/wg12>
Vid. IEEE
- MAB** Monitoring and Assessment of Biodiversity
<http://www.si.edu/simab/datamgt3.htm>
- MALIBU** MAnaging the hybrid LIbrary for the Benefit of Users
<http://www.kcl.ac.uk/humanities/cch/malibu>
- MARBI** Machine-Readable Bibliographic Information
<http://lcweb.loc.gov/marc/marbi.html>
- MARC** Machine Readable Cataloging
<http://www.loc.gov/marc>
- MBP** MetaBot Project Files (extensión de los archivos generados con Metabot)
- MCF** Meta Content Framework/Format
<http://www.w3.org/TR/NOTE-MCF-XML>
- MEG** Metadata Educational Group
<http://www.ukoln.ac.uk/metadata/education>

Sigla Significado/ URL

- MEMRI** Media Error Monitoring and Reporting Information
- MeSH** Medical Subject Headings
<http://www.nlm.nih.gov/mesh/meshhome.html>
- MESL** Museum Educational Site Licensing Project
<http://www.inform.umd.edu/EdRes/FacRes/TeachTech/.MESL>
- METACHEM** (Catalogue of chemistry resources)
<http://metachem.ch.adfa.edu.au>
- METAe** The Metadata Engine Project
<http://meta-e.uibk.ac.at>
- MIME** Multipurpose for Internet Mail Extensions
<http://www.ietf.org/rfc/rfc822.txt>
<http://www.ietf.org/rfc/rfc1521.txt>
Vid. IMT, RFC 822, RFC 1521
- MLA** Modern Language Association
<http://www.mla.org>
- MMI** Metadata for Multimedia Information
Vid. CEN/ISSS
- MMM** Mathematics Metadata Mark-up Editor
<http://www.mathematik.uni-osnabrueck.de/cgi-bin/MMM3.0.cgi>
- MNIS** Manning & Napier Information Services
<http://www.mnis.net>
- MOA** The Making Of America
<http://moa.umdl.umich.edu>
<http://moa.cit.cornell.edu/moa/index.html>
MOA II
<http://sunsite.berkeley.edu/moa2>
Vid. DLI
- MODELS** MOving to Distributed Environments for Library Services
<http://www.ukoln.ac.uk/dlis/models>
- MOF** Meta Object Facility Specification
Vid. OMG
- MOMIS** Mediator envirOnment for Multiple Information Sources
<http://sparc20.ing.unimo.it/Momis>
- MPEG** Moving Picture Experts Group
Normas del MPEG: <http://mpeg.telecomitalialab.com/standards.htm>
Vid. MPEG7
- MPEG7** Multimedia Content Description Interface
Vid. MPEG y DDL
- MSC** Mathematics Subject Classification
<http://www.ams.org/msc>

Índice de siglas y acrónimos

Sigla Significado/ URL

MSN	MicroSoft Network http://www.msn.com
MTP	Metadata Transaction Protocol (mtp://)- (macros de Metabrowser)
NACO	Name Authority Component http://www.loc.gov/catdir/pcc/naco.html <i>Vid. PCC</i>
NBII	National Biological Information Infrastructure http://www.nbii.gov
NCSA	National Center for Supercomputing Applications http://access.ncsa.uiuc.edu
NCSTRL	Networked Computer Science Technical Reference Library http://www.ncstrl.org
NDLTD	Networked Digital Library of Theses and Dissertations http://www.ndltd.org
NEDLIB	Networked European Deposit Library http://www.kb.nl/coop/nedlib
NEH	National Endowment for the Humanities http://www.neh.fed.us
NESSTAR	Networked Social Science Tools And Resources http://www.nesstar.org
NII	National Information Infrastructure
NISO	National Information Standards Organization http://www.niso.org http://www.techstreet.com/nisogate.html (normas NISO)
NIST	National Institute of Standards and Technology http://www.nist.gov
NKOS	Knowledge Organization Sources/Systems (NKOS) http://nkos.slis.kent.edu
NLA	National Library of Australia http://www.nla.gov.au
NLE	National Library of Education http://www.ed.gov/NLE
NLII	National Learning Infrastructure Initiative http://www.educause.edu/nlii
NOF	New Opportunities Fund http://www.nof.org.uk http://www.ukoln.ac.uk/nof/support/intro.htm

Sigla Significado/ URL

NSDI	National Spatial Data Infrastructure http://www.fgdc.gov/nsdi/nsdi.html
NSF	National Science Foundation http://www.nsf.gov
NZGLS	New Zealand Government Locator Service http://www.e-government.govt.nz/guidelines/nzgl
NZGP	New Zealand Government Portal http://www.govt.nz
OAI	Open Archive Initiative http://www.openarchives.org OAI protocol http://www.openarchives.org/OAI_protocol/openarchivesprotocol.html
OAIS	Open Archival Information System http://www.ccsds.org/RP9905/RP9905.html http://www.ccsds.org/documents/pdf/CCSDS-650.0-R-1.pdf
OASIS	Organization for the Advancement of Structured Information Standards http://www.oasis-open.org
Object ID	Estándar internacional para la descripción de obras de arte, antigüedades y artefactos arqueológicos http://www.object-id.com
OCLC	Online Computer Library Center http://www.oclc.org/home
OCR	Optical Character Recognition
OCS	Open Content Syndication http://internetalchemy.org/ocs/directory.html <i>Vid. RSS</i>
ODBC	Open Database Connectivity
ODP	Open Directory Project http://dmoz.org
ODRL	Open Digital Rights Language http://odrl.net
OEI	Organización de Estados Iberoamericanos http://www.oei.org Biblioteca Digital de la OEI: http://www.campus-oei.org/oeivirt
OGC	OpenGIS Consortium http://www.opengis.org <i>Vid. GIS, SIG.</i>
OII	Open Information Interchange http://www.diffuse.org/oii/en/oii-home.html

Sigla Significado/ URL

- OIL** Ontology Inference Layer/Ontology Interchange Language
<http://www.ontoknowledge.org/oil>
- OLAP** Online Analytical Processing
<http://www.olapcouncil.org>
- OLSTF** [Formato propietario de NewsAgent –eLib–]
- OMG** Object Management Group
<http://www.omg.org>
- OMNI** Organising Medical Networked Information
<http://omni.ac.uk>
- ONIX** Online Information eXchange
<http://www.publishers.org/onix.htm>
- OPAC** Online Public Access Catalog
Vid WebPAC
- OpenURL** Open Uniform Resource Locator
<http://www.sfxit.com/openurl>
<http://www.sfxit.com/openurl/openurl.html>
- ORB** Object Request Broker
<http://www.sei.cmu.edu/str/descriptions/orb.html>
Vid. CORBA
- OTA** Open Text Archive
<http://ota.ahds.ac.uk>
Vid. AHDS
- PAC** Policy Advisory Committee
Vid. DC, DCMI
- PACS** Physics and Astronomy Classification Scheme
<http://publish.aps.org/PACS/pacsgen.html>
- PADI** Preserving Access to Digital Information
<http://www.nla.gov.au/padi>
- PANDORA** Preserving and Accessing Networked Documentary Resources in Australia
<http://pandora.nla.gov.au>
- PAS** Public Available Specifications
- PCC** Program for Cooperative Cataloging
<http://www.loc.gov/catdir/pcc>
<http://www.loc.gov/catdir/pcc/pccinfsp.html> (en español)
Vid. BIBCO, CONSER, NACO, SACO
- PDF** Portable Document Format (formato de archivo propietario de Adobe)

Sigla Significado/ URL

PICS	Platform for Internet Content Selection http://www.w3.org/PICS
PINAKES	A Subject Launchpad http://www.hw.ac.uk/libWWW/irn/pinakes/pinakes.html
PLN	Procesamiento del Lenguaje Natural
PNG	Portable Network Graphics: http://www.libpng.org/pub/png http://www.w3.org/TR/REC-png
PR	Proposed Recommendation (W3C)
PRISM	Publishing Requirements for Industry Standard Metadata http://prismstandard.org http://www.prismstandard.org/techdev/prismspec1.asp PRISM es también el nombre con el que se conoce al proyecto de bibliotecas digitales de la Universidad de Cornell denominado: Information Integrity in Distributed Digital Libraries http://www.prism.cornell.edu <i>Vid. DLI2</i>
PS	PostScript (extensión de fichero/tipo de archivo de ordenador)
PURL	Persistent Uniform Resource Locator http://purl.oclc.org
QGMS	Queensland Government Metadata IS#34 http://www.slq.qld.gov.au/meta/guidelines.htm#qgms <i>Vid. QLDGOV</i>
QLDGOV	Queensland Government http://www.qld.gov.au <i>Vid. QGMS</i>
RC	Reglas de Catalogación Españolas <i>Vid. RRCC</i>
RDBMS	Relational Data Base Management System
RDDL	Resource Directory Description Language http://www.openhealth.org/RDDL
RDF	Resource Description Framework http://www.w3.org/RDF
RDFMS	RDF Model and Syntax Specification http://www.w3.org/TR/REC-rdf-syntax
RDFS	RDF Schema http://www.w3.org/TR/rdf-schema

Sigla Significado/ URL

- RDN** Resource Discovery Network
<http://www.rdn.ac.uk>
- RDU** Resource Discovery Unit.
<http://www.dstc.edu.au/RDU>
- REC** Recommendation (W3C)
- Renardus** Academic Subject Gateway Service Across Europe
<http://www.renardus.org>
<http://www.ukoln.ac.uk/metadata/renardus>
- REP** Robots Exclusion Protocol
- RFC** Request For Comments
<http://www.ietf.org/rfc.html>
<http://www.nexor.com/index-rfc.htm>
<ftp://ftp.isi.edu/in-notes>
- RFC 822:** Standard for ARPA Internet Text Messages
<http://www.ietf.org/rfc/rfc822.txt>
- RFC 1521:** Multipurpose Internet Mail Extensions (MIME)
<http://www.ietf.org/rfc/rfc1521.txt>
- RFC 1766:** Tags for the Identification of Languages (ISO639). <http://www.ietf.org/rfc/rfc1766.txt>
- RFC 1807:** A Format for Bibliographic Records
<http://www.ietf.org/rfc/rfc1807.txt>
- RFC 1835:** Architecture of the WHOIS++ service
- RFC 2376:** XML Media Types (*Vid.* RFC 3023)
<http://www.ietf.org/rfc/rfc2376.txt>
- RFC 2396:** Uniform Resources Identifiers (URI): Generic syntax. <http://www.ietf.org/rfc/rfc2396.txt>
- RFC 2413:** Dublin Core Metadata for Resource Discovery
<http://www.ietf.org/rfc/rfc2413.txt>
- RFC 2616:** Hypertext Transfer Protocol
<http://www.ietf.org/rfc/rfc2616.txt>
- RFC 2651** Common Indexing Protocol
<http://www.ietf.org/rfc/rfc2651.txt>
- RFC 2731:** Encoding Dublin Core Metadata in HTML
<http://www.ietf.org/rfc/rfc2731.txt>
- RFC 3023:** XML Media Types (*Vid.* RFC 2376)
<http://www.ietf.org/rfc/rfc3023.txt>
- Request For Comments en español: <http://www.rfc-es.org>
Vid. IETF
- RFI** Request for Information
- RI** Recuperación de Información
Vid. IR, SRII

Sigla Significado/ URL

- RIGILS** Rhode Island Government Locator Service
<http://www.gils.state.ri.us/pg1.htm>
Vid. GILS
- RLG** Research Library Group
<http://www.rlg.org/toc.html>
- ROADS** Resource Organisation And Discovery in Subject-based Services
<http://www.ilrt.bris.ac.uk/roads>
<http://www.ukoln.ac.uk/metadata/roads>
- ROI** Registration Of Interest
- RRCC** Reglas de Catalogación (sentido genérico)
Vid. RC
- RSAC** Recreational Software Advisory Council
Vid. ICRA
- RSLP** Research Support Library Program
<http://www.rslp.ac.uk>
<http://www.ukoln.ac.uk/metadata/rslp>
- RSS** RDF Site Summary (antes, Rich Site Summary)
<http://groups.yahoo.com/group/rss-dev/files/specification.html>
<http://www.24horas.com.pe/tutrss.php>
- SAB** Sistema de encabezamientos de material y códigos de clasificación de Suecia: <http://www.amnesord.kb.se>
- SACO** Subject Authority Component
<http://www.loc.gov/catdir/pcc/saco.html>
Vid. PCC
- SAFARI** Base de datos sobre investigación (Suecia).
(El acrónimo en sueco significa: la divulgación de la información de investigación al público general a través de Internet)
<http://safari.vr.se>
- SARAC** Servicio de Acceso a Recursos de Alta de Calidad
- SBIGs** Subject-Based Information Gateways
- SCHEMAS** Forum for Metadata Schemas Implementers
<http://www.schemas-forum.org>
- SDD** Standalone Document Declaration
Vid. XML, DTD.
- SDI** Single Document Interface
- SDK** Software Development Kits (MetaStar)
<http://www.blueangelttech.com/products/Components/SDKs.htm>

Sigla Significado/ URL

SEDIC	Sociedad Española de Documentación e Información Científica http://www.sedic.es
SFX	Tecnología patentada por Ex-Libris: Context-sensitive reference linking http://www.sfxit.com
SGBD	Sistema de Gestión de Bases de Datos <i>Vid.</i> DBMS
SGBDR	Sistema de Gestión de Bases de Datos Relacionales <i>Vid.</i> RDBMS
SGML	Standard Generalized Markup Language http://www.oasis-open.org/cover/sgml-xml.html
SH	Semantic Header http://www.cs.concordia.ca/~faculty/bcdesai/cindi-system-1.0.html
SHML	Structured Hypertext Markup Language
SHOE	Simple HTML Ontology Extensions http://www.cs.umd.edu/projects/plus/SHOE
SICI	Serial Item and Contribution Identifier http://sunsite.berkeley.edu/SICI <i>Vid.</i> Z39.56
SIG	Special Interest Group Sistemas de Información Geográfica. <i>Vid.</i> GIS
SIGB	Sistema Integrado de Gestión de Bibliotecas
SIGBD	Sistema Integrado de Gestión de Bibliotecas Digitales <i>Vid.</i> SIG-DLOs
SIG-DLOs	Sistema Integrado de Gestión de DLOs <i>Vid.</i> DLO
SlinkS	Scholarly Link Specification http://www.openly.com/SLinkS/SLinkS.html http://www.openly.com/SLinkS/Editor.html (software)
SMIL	Synchronized Multimedia Integration Language http://www.w3.org/AudioVideo
SMPTE	Society of Motion Picture and Television Engineers http://www.smpte.org
SOAP	Simple Object Access Protocol http://www.w3.org/TR/SOAP http://www.develop.com/soap
SOIF	Summary Object Interchange Format http://www.ukoln.ac.uk/metadata/desire/overview/rev_20.htm

Sigla Significado/ URL

SOSIG	Social Science Information Gateway http://sosig.ac.uk
SOX	Schema for Object-Oriented XML http://www.w3.org/TR/NOTE-SOX
SPECTRUM	The UK Museum Documentation Standard http://www.mda.org.uk/spectrum.htm
SQL	Standard Query Language
SRII	Sistemas de Recuperación de Información en Internet
SRW	Search/Retrieve Web Service http://www.loc.gov/z3950/agency/zing/srw.html
STAR	Strathclyde Authored Resources http://wp269.lib.strath.ac.uk:5050/Cat2/demo_guide.html
STARTS	Stanford Protocol Proposal for Internet Search and Retrieval http://www-db.stanford.edu/~gravano/starts_home.html
SUTRS	Simple Unstructured Text Record Syntax http://www.loc.gov/z3950/agency/clarify/sutrs.html
SW	Semantic Web http://www.semanticweb.org http://www.w3.org/RDF <i>Vid. WS, RDF</i>
SWAG	Semantic Web Agreement Group http://swag.webns.net
TAC	Technical Advisory Committee <i>Vid. DC</i>
TACT	Text Analysis Computing Tools http://www.chass.utoronto.ca/cch/tact.html
TC46/SC4	Technical Committee 46 Subcommittee 4 TC 46 - Information and Documentation SC 4 - Computer Applications http://www.niso.org/tc46sc4.html http://www.iso.ch/iso/en/stdsdevelopment/tc/tclist/TechnicalCommitteeDetailPage.TechnicalCommitteeDetail?COMMID=1778 <i>Vid. ISO</i>
TCP/IP	Transmission Control Protocol / Internet Protocol
TDDL	Terminology Definition Description Language http://www.openhealth.org/RDDL/tddl <i>Vid. RDDL</i>

Sigla Significado/ URL

- TEI** Text Encoding Initiative
<http://www.tei-c.org>
<http://www.hcu.ox.ac.uk/TEI> (mirror)
- TEI Lite** Text Encoding Initiative Lite (Encoding for Interchange: an introduction to the TEI)
<http://www.tei-c.org/Lite/index.html>
<http://www.hcu.ox.ac.uk/TEI/Lite>
Vid. TEI U5
- TEI P3** Guidelines for Electronic Text Encoding and Interchange
<http://www.hti.umich.edu/t/tei>
<http://etext.virginia.edu/TEI.html>
- TEI P4** Text Encoding Initiative (XML, Junio 2001)
<http://www.tei-c.org/P4X/DTD>
- TEI U5** *Vid.* TEI Lite
- TEI-C** TEI Consortium
<http://www.tei-c.org>
<http://www.hcu.ox.ac.uk/TEI> (mirror)
- TEIH** TEI Header
<http://www.tei-c.org/P4X/HD.html>
<http://www.tei-c.org/P4X/DTD/teihdr2.dtd>
- TEL** The European Library
<http://www.europeanlibrary.org>
- TGN** Thesaurus of Geographical Names
<http://www.getty.edu/research/tools/vocabulary/tgn>
- TI** Tecnología/s de Información
- TIC** Tecnología de la Información y de las Comunicaciones.
(También TICs, en plural)
- TIFF** Tagged Image File Format
- TR** Technical Report
Vid. OASIS, W3C
- TRS** Terminology Reference System
Vid. EPA
- TRAIL** Texas Records and Information Locator Service
<http://castor.tsl.state.tx.us/trail/index.html>
- UDC** Universal Decimal Classification
<http://www.udcc.org>
- UKOLN** United Kingdom Office for Library and Information Networking
(Universidad de Bath)
<http://www.ukoln.ac.uk>

Sigla Significado/ URL

- ULAN** Union List of Artist Names
<http://www.getty.edu/research/tools/vocabulary/ulan>
- UMDL** University of Michigan Digital Library
<http://www.si.umich.edu/UMDL>
Vid. DLI
- UML** Unified Modeling Language
<http://www.rational.com/uml/resources>
UML (v. 1.3)
<http://www.rational.com/uml/resources/documentation/index.jsp>
- UNE** Norma española
<http://www.aenor.es/scripts/buscar.asp?idioma=E>
- UNEP** United Nations Environment Programme
<http://www.unep.org>
<http://www.unep.net>
- URC** Uniform Resource Characteristics/Citations
http://www.ukoln.ac.uk/metadata/desire/overview/rev_22.htm
- URI** Uniform Resource Identifier
<http://www.ietf.org/rfc/rfc2396.txt>
<http://www.w3.org/Addressing>
Vid. RFC 2396
- URL** Uniform Resource Locator
<http://www.w3.org/Addressing>
- URN** Uniform Resource Name
<http://www.ietf.org/html.charters/urn-charter.html>
<http://www.w3.org/Addressing>
- USGS** United States Geological Survey
<http://www.usgs.gov>
- VADS** Visual Arts Data Service
<http://vads.ahds.ac.uk>
<http://vads.ahds.ac.uk/search.html>
Vid. AHDS
- VALA** Victorian Association for Library Automation
<http://www.vala.org.au>
- VGIS** Vermont Geographic Information System
<http://geo-vt.uvm.edu>
<http://geo-vt.uvm.edu/cfdev2/Library/meta/ccap.cfm> (metadatos)
- VIA** Visual Information Access
<http://hul.harvard.edu/ois/systems/via>
- VINIS** Virtual Library and Information System
<http://mgu.vinis.co.kr>
- VISION** Video Indexing for Searching Over Networks

Sigla Significado/ URL

VLDB	Very Large Data Bases http://www.vldb.org
VOC-ML	Vocabulary Markup Language http://nkos.slis.kent.edu/VOCML-1.DOC <i>Vid. NKOS</i>
VRA	Visual Resources Association http://www.vraweb.org VRA Core Categories v. 3 http://php.indiana.edu/~fryp/vracore3.htm http://www.gsd.harvard.edu/%7Estaffaw3/vra/vracore3.htm
VRML	Virtual Reality Modeling Language (ISO/IEC 14772) http://vrml.org http://vrml.org/VRML1.0/vrml10c.html
VTC	Virtual Teacher Centre Metadata Schema http://vtc.ngfl.gov.uk/docserver.php?docid=2190
W3C	World Wide Web Consortium http://www.w3.org
W3CDTF	W3C Data and Time Format (ISO 8601) http://www.w3.org/TR/NOTE-datetime
W3D	Web 3D Consortium http://www.web3d.org
WAGILS	Washington Government Information Locator Service <i>Vid. GILS</i>
WAI	Web Accessibility Initiative http://www.w3.org/WAI
WCAG	Web Content Accessibility Guidelines (1.0) http://www.w3.org/TR/WCAG10 <i>Vid. WAI</i>
WD	Working Draft (W3C)
WEBDAV	World Wide Web Distributed Authoring and Versioning http://www.webdav.org http://www.ics.uci.edu/pub/ietf/webdav
WebPAC	Web Public Access Catalog
WebOnt	Web Ontology Working Group http://www.w3.org/2001/sw/WebOnt
WebOPAC	<i>Vid. WebPAC (se sobreentiende que el acceso Web es Online)</i>
WHOIS++	Protocolo de los servicios de directorio en Internet http://www.ietf.org/rfc/rfc1835.txt <i>Vid. RFC 1835</i>

Sigla Significado/ URL

WISCLIC	Wisconsin Land Information Clearinghouse http://wisclinc.state.wi.us
WODA	Web Oriented Database http://www.ddatabase.com
WP3	Work Package 3 (Grupo de trabajo del proyecto DESIRE).
WS	WebSemantics http://www.cs.toronto.edu/~georgem/ws <i>Vid.</i> SW, RDF
WSDL	Web Service Description Language http://www.w3.org/TR/wsd http://xml.coverpages.org/wsd.html
WSE	Web Search Environments (ILRT) http://wse.search.ac.uk
WSSN	Web Standards Service Network http://www.wssn.net/WSSN
WWW	World Wide Web
WWW VL	World Wide Web Virtual Library http://vlib.org
WYSIWYG	What You See Is What You Get
X3D	Extensible 3D http://www.web3d.org/x3d
XER	XML Encoding Rules http://www.dstc.edu.au/Research/Projects/xer
XHTML	eXtensible HyperText Markup Language http://www.w3.org/MarkUp http://www.xhtml.org
XLink	XML Linking Language http://www.w3.org/XML/Linking
XLL	<i>Vid.</i> Xlink
XMI	eXtensible Metadata Interchange http://www.oasis-open.org/cover/xmi.html
XML	eXtensible Markup Language http://www.w3.org/XML
XML:DB	XML: Data Bases http://www.xmldb.org

Sigla Significado/ URL

xmlns	XML Namespaces http://xmlns.com
XMLP	XML Packages (MetaPackager, TagGen Office)
XMP	eXtensible Metadata Platform (Antes XAP: Iniciativa de Adobe para incluir metadatos en el fichero binario de los documentos pdf) http://www.adobe.com/products/xmp/main.html
XOL	Ontology Exchange Language http://www.ai.sri.com/~pkarp/xol
XP	XML Protocol http://www.w3.org/2000/xp
XPath	XML Path Language http://www.w3.org/TR/xpath
XQuery	XML Query http://www.w3.org/XML/Query
XSL	eXtensible Style Sheet Language http://www.w3.org/Style/XSL
XSLT	XSL Transformations http://www.w3.org/TR/xslt <i>Vid. XSL</i>
XTM	XML Topic Maps http://www.topicmaps.org
XUA	XML User Agent
Z39.50	Norma ANSI/NISO Z39.50-1995 (ISO23950-1998) Information Retrieval: Application Service Definition & Protocol Specification http://www.techstreet.com/cgi-bin/pdf/free/228859/z39-50a.pdf http://www.techstreet.com/cgi-bin/pdf/free/228860/z39-50b.pdf http://www.techstreet.com/cgi-bin/pdf/free/228861/z39-50c.pdf http://www.techstreet.com/cgi-bin/pdf/free/228862/z39-50d.pdf http://www.techstreet.com/cgi-bin/pdf/free/228863/z39-50e.pdf Revisión de la norma Z39-50- 2001 : http://www.loc.gov/z3950/agency/revision/part1.pdf http://www.loc.gov/z3950/agency/revision/part2.pdf
Z39.56	Norma ANSI/NISO Z39.56-1996 Serial Item & Contribution Identifier (SICI) http://www.techstreet.com/cgi-bin/pdf/free/152629/z39-56.pdf <i>Vid. SICI</i>
Z39.84	Norma ANSI/NISO Z39.84-2000 Syntax for the Digital Object Identifier http://www.techstreet.com/cgi-bin/pdf/free/247384/z39-84.pdf <i>Vid. DOI</i>

Sigla Significado/ URL

- Z39.85** Norma ANSI/NISO Z39.85-2001
The Dublin Core Metadata Element Set
<http://www.niso.org/pdfs/Z39-85.pdf> (agosto 2000)
<http://www.niso.org/pdfs/Z3985-REV.pdf> (marzo 2001)
<http://www.niso.org/standards/resources/Z39-85.pdf> (septiembre 2001)
Vid. DCMI, DCMES
- ZDM** Zentralblatt für Didaktik der Mathematik
International Reviews on Mathematical Education
<http://www.mathematik.uni-osnabrueck.de/projects/zdm>
- ZETOC** Z39.50 Electronic Table of Contents
(Electronic Table of Contents from the British Library)
<http://zetoc.mimas.ac.uk>
- ZIG** Z39.50 Implementors (sic.) Group
<http://www.loc.gov/z3950/agency/zig/zig.html>
- ZING** Z39.50 International Next Generation
<http://www.loc.gov/z3950/agency/zing/zing.html>
- ZML** Z39.50 Markup Language
- ZNG** Z39.50 Next Generation
Vid. ZING
- ZOOM** Z39.50 Object-Oriented Model
<http://zoom.z3950.org>