



ÍNDICE

INTRODUCCIÓN	7
Objetivos	8
Marco teórico	9
Metodología	12
Recursos electrónicos	14
<i>Tipología</i>	15
CAPÍTULO I. CATALOGACIÓN DE LOS RECURSOS ELECTRÓNICOS	19
CAPÍTULO II. METADATOS	23
Funciones	25
Tipología	27
CAPÍTULO III. NORMATIVAS INTERNACIONALES PARA LA DESCRIPCIÓN DE RECURSOS ELECTRÓNICOS	29
Estándares de contenido	30
<i>AACR2R</i>	30
<i>ISBD</i>	32
CAPÍTULO IV. ANÁLISIS DE LOS ESQUEMAS DE METADATOS	35
MARC 21/MARCXML	37
Dublin Core	45
MODS (Metadata Object Description Schema)	60
TEI (Text Encoding Initiative)	79

CAPÍTULO V. LENGUAJES DE MARCADO	91
SGML	91
HTML	92
XML	93
RDF	94
CAPÍTULO VI. COMPARACIÓN DE LOS ESQUEMAS DE METADATOS	97
Estructura, contenido, sintaxis	99
Simplicidad, extensibilidad, flexibilidad	100
Interoperabilidad	101
CAPÍTULO VII. METADATOS Y RECUPERACIÓN DE LA INFORMACIÓN	103
Buscadores o motores de búsqueda	104
Metabuscadore s (<i>metasearch engines</i>)	106
Directorios (<i>web directories</i>)	107
CONCLUSIONES	109
BIBLIOGRAFÍA CONSULTADA	113



INTRODUCCIÓN

La sociedad de la información y la globalización conllevan a un cambio de paradigma en el campo de la organización del conocimiento. La información electrónica presenta una arquitectura hipertextual, no lineal y distribuida que plantea nuevos desafíos al profesional bibliotecario.

Internet ha permitido el acceso remoto a los catálogos y a los instrumentos de descripción, e incluso los propios fondos documentales de las bibliotecas están siendo digitalizados y puestos a disposición de los usuarios a través de este poderoso medio de comunicación. Una eclosión de diversos tipos de documentos va poblando Internet a un ritmo de crecimiento exponencial. Una consecuencia de ello es que cualquier catálogo de biblioteca puede abastecerse de la gran cantidad de documentos que se distribuyen por Internet, superando los fondos locales y abriéndose al inmenso espacio digital. Esto está claro en la teoría, pero aún muchas

bibliotecas no catalogan fondos de Internet, aunque sí ofrezcan un espacio de consulta y un servicio de referencia virtual.

Como lo plantean Pinto Molina, García Marco y Agustín Lacruz (2002), si bien no es una experiencia nueva el hecho de que una parte muy importante de los documentos de interés para los usuarios esté fuera de la propia biblioteca, sí es una situación innovadora que comiencen a ser extraordinariamente accesibles. Más adelante, estos autores, que señalan la facilidad con la que se pueden digitalizar los documentos textuales y audiovisuales, la transmisión instantánea a distancia, su simultaneidad de uso, su utilización sin deterioro y la reducción del espacio de almacenamiento, ponen de manifiesto que con los documentos digitales han desaparecido las limitaciones espaciales y temporales intrínsecas al resto de los documentos, lo que dispara de un modo espectacular sus posibilidades informativas. Por ello, la integración de registros bibliográficos de materiales remotos se ha convertido en una creciente necesidad para la creación y el mantenimiento de los catálogos de las bibliotecas.

En este estudio se aborda el tema de los metadatos y las normativas internacionales para la descripción de recursos electrónicos de información disponibles en línea, analizando los problemas que comportan a nivel conceptual y de desarrollo de nuevas formas de estructurar la información. Desde una perspectiva bibliotecológica, se da preponderancia a la normalización y a las posibilidades de aplicación de los desarrollos estudiados para organizar la información accesible vía Web.

La información digital difiere de la documentación impresa en muchos aspectos, por lo que se ha generado un debate en el ámbito de los profesionales bibliotecarios con respecto a los métodos tradicionales de procesamiento de la información. En este trabajo se trata de responder preguntas tales como: ¿Qué son realmente los metadatos? ¿Es un concepto realmente nuevo? ¿Cuál es la diferencia con la catalogación tradicional? ¿Para qué sirven? ¿Qué normativas han surgido? ¿Qué diferencias y similitudes hay entre ellas?

Objetivos

- Analizar similitudes y diferencias entre la catalogación de materiales tradicionales y de recursos de información en Internet.
- Analizar diferentes esquemas de metadatos para la descripción de recursos electrónicos accesibles por Internet.

- Comparar los esquemas analizados sobre la base de los criterios establecidos.
- Elaborar conclusiones en cuanto a los resultados de la comparación y en relación con la recuperación de información en la Web.

Marco teórico

Los metadatos tienen un rol importante en el contexto del surgimiento de las bibliotecas digitales, ya que la organización de la información disponible en las redes depende de la administración efectiva de ellos.

En este trabajo se entiende por metadatos, siguiendo a Taylor (2004), a la descripción codificada de un paquete de información cuyo propósito es proporcionar un nivel intermedio a través del cual se puede acceder a la información por medio de lo que uno desea ver o buscar, sin tener que investigar enormes cantidades de texto completo irrelevante.

Burnett, Bor Ng y Park (1999) señalan que los metadatos han sido abordados desde dos ópticas:

- **la del control bibliográfico**, con origen y proponentes desde la bibliotecología; y
- **la del procesamiento de datos**, con origen y proponentes desde las ciencias de la computación.

A pesar de las diferencias, estos dos abordajes usan metadatos para localizar, identificar, recuperar y administrar la información; y a medida que avanza el desarrollo, estas dos corrientes se van acercando e integrando.

Eito Brum (2008) manifiesta que la proliferación de los documentos publicados en la Web hizo que la comunidad bibliotecaria aplicara técnicas similares a la catalogación tradicional para la identificación, localización y posterior recuperación de estos nuevos recursos.

Las bibliotecas tienen una larga tradición en el desarrollo de sistemas de información, y se concentran en organizar y proveer acceso a los documentos. Para ello utilizan la descripción bibliográfica, el análisis de contenido y la clasificación, de modo que cada entidad de información pueda ser eficientemente descripta, identificada, localizada y recuperada. Mediante el establecimiento de una normativa específica, han logrado implementar con éxito complejos sistemas documentales. Así, la unidad básica del catálogo, el registro bibliográfico, representa unívocamente y describe de forma precisa un documento.

La necesidad de intercambiar información entre instituciones a nivel local, nacional e internacional ha hecho que se desarrollen normas y procedimientos que aseguren la interoperabilidad de los sistemas de información. Aunque no se resuelvan todos los requerimientos de la organización de la información digital, las reglas que se vienen usando sirven de base y tienen un impacto significativo.

Las ciencias de la computación también tienen un largo camino recorrido en cuanto al procesamiento de datos y a la organización de la información. Hay numerosos sistemas computarizados de almacenamiento y recuperación de información, tanto para datos textuales como relacionales, que proveen seguridad, integridad y posibilidad de intercambio de datos, además de las funciones tradicionales de localización, identificación y recuperación. Todo el avance logrado desde las ciencias de la computación también tiene una fuerte influencia en el desarrollo de metadatos.

Desde la perspectiva del control bibliográfico, se desarrollaron metadatos orientados a la catalogación de los recursos de información en Internet, como la utilización del formato MARC (*MACHine Readable Cataloguing*) como estructura de metadatos y la formulación de Dublin Core. Esta corriente se focaliza en la descripción e identificación de los objetos de información.

Desde la perspectiva de la computación, se dio preponderancia a la manipulación de los datos para su uso, más allá de su mera descripción, y se incluyeron datos de acceso, de autenticación, administrativos, etc. Se han desarrollado lenguajes de marcación de textos como HTML (*Hyper Text Mark-up Language*), XML (*eXtensible Mark-up Language*), SGML (*Standard Generalized Mark-up Language*), Warwick Framework, y formatos como IAFA (*Internet Anonymous FTP Archives*) y TEI (*Text Encoding Initiative*).

A pesar de que estas dos aproximaciones a los metadatos son diferentes, de ningún modo son excluyentes o incompatibles, más bien son complementarias, y la solución definitiva provendrá de un estudio interdisciplinario, que integre técnicas de una y otra.

Como señala Taylor (2004), entre los profesionales de la información el concepto de metadatos puede aparecer complejo y confuso. Esto se debe, en parte, a la naturaleza multifacética del tema. Es importante recordar que el término metadatos puede significar diferentes cosas para diferentes comunidades. Cuando un bibliotecario habla de metadatos, tiene una noción diferente de la que posee quien trabaja, por ejemplo, con metadatos geoespaciales.

Dovey (1999) identifica tres corrientes principales a las que denomina escuelas: “*la escuela de la catalogación*”, “*la escuela estructuralista*” y “*la es-*

cuela de la estructura de datos". La primera incluye a los catalogadores que proceden del mundo bibliotecario y que han estado generando metadatos prácticamente desde que existen las bibliotecas; también aquí ubica a los informáticos preocupados por generar estructuras que hagan la información más manejable, y dedicados a construir metadatos para catalogar. En esta línea se ubican los creadores de Dublin Core y RDF. La segunda, la escuela estructuralista, está constituida por los creadores de los lenguajes de marcado, cuyos estándares y modelos están basados en la estructura del texto, como por ejemplo HTML o SGML. La tercera de estas escuelas, llamada de estructura de datos, pone de relieve el problema de la separación del contenido de la presentación en los documentos electrónicos; esta corriente ve al XML como un lenguaje universal para describir estructuras de datos, adecuado para el intercambio tanto de la estructura de las bases de datos como del contenido.

Tal como lo señalan Pinto Molina, García Marco y Agustín Lacruz (2002), si algo caracteriza el momento actual es la superación de la ecuación tradicional "documento = texto verbal escrito". La aparición de diferentes soportes a lo largo del tiempo ha llevado a utilizar ciertos nuevos términos muy imprecisos como "materiales especiales", "nuevos documentos", "documentos no textuales". El problema se complica aun más con los documentos multimedia y virtuales. Una de las características de éstos es que integran los tipos anteriores, por ejemplo un multimedia interactivo contiene textos escritos, sonidos, imágenes y videos.

En la actualidad, cualquier documento se puede convertir a formato digital; así cuando se habla de documento digital o electrónico, se hace referencia al formato de codificación utilizado para su almacenamiento y procesamiento más que a la forma de presentación de la información contenida en él. Cuando se utiliza el término documento digital sólo para aquéllos distribuidos por Internet, se confunde formato y medio, ya que Internet es un medio de comunicación y dichos documentos preceden a Internet.

La utilización del término "recurso de información" es de procedencia anglosajona donde se ha popularizado el término "*information resources*". En este trabajo se lo utiliza, si bien puede objetársele el ser poco específico, porque supera, en cierto modo, la dificultad terminológica aún no resuelta en cuanto a la tipología documental. Es importante destacar que las ISBD(CF) (*International Standard Bibliographic Description: Computer Files*) se revisan en 1997, y se cambia la denominación de "archivos de ordenador" por "recursos electrónicos", dando origen a la ISBD(ER) (*International Standard Bibliographic Description: Electronic Resources*).

Siguiendo esta línea, otro concepto que se adopta en este trabajo es el de "objeto de información", DLO (*Document Like Object*) según su sigla en

inglés. Tal como lo plantea Méndez Rodríguez (2002), una de las claves principales de la teoría de metadatos está en definir qué es un objeto de datos o qué es un documento entendido como objeto. Así, en un ambiente digital, se puede definir un DLO como un elemento o un grupo de elementos que constituyen una unidad informativa que una computadora puede manejar, independientemente del formato del archivo y del tipo de información que reflejen. Entendido así, en las páginas subsiguientes se utilizan términos como recurso electrónico, documento electrónico, objeto digital u objeto de información, para referirse a esta entidad.

Metodología

El presente trabajo responde a un diseño bibliográfico o investigación documental. Tamayo y Tamayo (2000) definen la investigación documental como “la que se realiza con base a la revisión de documentos, manuales, revistas, periódicos, actas científicas, conclusiones de simposios y seminarios y/o cualquier tipo de publicación considerado como fuente de información”. El presente estudio se limita a aquellos proyectos cuyos responsables son instituciones de reconocido prestigio en el campo de la bibliotecología, tales como la *Library of Congress* (LC) u *Online Computer Library Center* (OCLC), a emprendimientos cooperativos internacionales y de instituciones pioneras en redes digitales. Así, los esquemas de metadatos seleccionados para la comparación son: MARC 21/MARCXML, Dublin Core, MODS (*Metadata Object Description Schema*) y TEI (*Text Encoding Initiative*).

Se realizó una búsqueda sobre el tema en libros y publicaciones periódicas, y especialmente en Internet, ya que la mayor parte de la información relevante y actualizada sobre la Web está en la misma Web. Se analizó bibliografía publicada a partir de 1994, año en el que el tema de metadatos cobra fuerza en el ámbito bibliotecológico y llega hasta la fecha de cierre de este trabajo, en diciembre de 2009. Para la descripción de los elementos y características de los esquemas de metadatos se utilizaron las fuentes primarias de información, es decir los sitios web oficiales de cada iniciativa.

La grilla de comparación se construyó sobre la base de los criterios propuestos por Heery (1996), Dempsey y Heery (1998) y Taylor (2004), y quedó conformada de la siguiente manera:

- **Disciplina/ámbito de aplicación:** considera quiénes están usando activamente este formato y si éste está asociado con una disciplina académica o interés profesional en particular.
- **Facilidad de creación de la descripción:** se refiere a la complejidad de la creación del registro, si se requieren habilidades especiales, si están diseñados para que asigne los metadatos el autor, el editor o los proveedores centralizados.
- **Contenido:** incluye estructura, sintaxis y contenido.
- **Protocolos de Internet asociados:** hace referencia a si los metadatos son soportados por los protocolos de Internet existentes y si una base de datos con dichos metadatos puede ser interrogada usando los protocolos existentes.
- **Contribución a la normalización internacional:** considera la ratificación del esquema de metadatos como norma internacional.
- **Interoperabilidad:** se refiere a la compatibilidad de dos o más sistemas de modo que puedan intercambiar datos e información y utilizar la información sin ninguna manipulación especial.
- **Extensibilidad:** tiene en cuenta la posibilidad de incluir mayor nivel de detalle, si fuera necesario, mediante campos opcionales, calificadores, etc.
- **Flexibilidad:** se puede aplicar a dos aspectos: a la flexibilidad para describir diferentes tipos de documentos y a la flexibilidad para evolucionar de acuerdo con los cambios en las reglas de catalogación y tecnología.

A través de este cotejo se pretende establecer cuáles son los metadatos que se perfilan con mayor posibilidad de éxito y difusión en el ámbito de la catalogación de recursos en red.

Recursos electrónicos

Lyons (1997) define un objeto digital o recurso electrónico como un conjunto de bits arreglados en una secuencia, que contiene datos que pueden ser interpretados, y que incluye un identificador único.

Según las Reglas de Catalogación Angloamericanas (segunda edición, actualización 2003, AACR2R), un recurso electrónico es un documento codificado para ser manipulado con la computadora. Se los divide, por un lado, en **datos** (información en la forma de números, letras, gráficos, imágenes y sonido, o combinación de cualesquiera de ellas) y **programas** (instrucciones o rutinas para ejecutar ciertas tareas); y por otro, se distinguen entre los de **acceso local** (con soporte físico, disco, casete, disquete) y **remoto** (disponibles a través de redes).

Según las normas de la Descripción Bibliográfica Internacional Normalizada (ISBD); (edición consolidada, 2007), los recursos electrónicos son materiales controlados por computadora, incluidos los que requieran el uso de periféricos (por ejemplo un lector de CD-ROM), y dichos recursos pueden ser o no utilizados en un modo interactivo. Los divide en: **datos** (información en la forma de números, letras, gráficos, imágenes y sonido, o combinación de cualesquiera de ellas) y **programas** (instrucciones o rutinas para ejecutar ciertas tareas); pueden ser también una **combinación de datos y programas** (por ejemplo, un software educativo que contenga texto, gráficos y programas).

Esta definición se aplica a los recursos que están disponibles e incluye también aquellos de acceso remoto vía telecomunicaciones. Sin embargo, los juguetes programados, las calculadoras y otros objetos programados se consideran fuera del alcance de la ISBD.

Una característica interesante de destacar es la cantidad de actores que intervienen en la producción y conservación de los recursos electrónicos (Lazinger, 2001):

- **Creadores:** individuos o instituciones generadores de la información contenida en ellos.
- **Editores:** en el caso de recursos publicados que poseen los derechos y licencias.
- **Distribuidores:** proveedores del acceso que no crean ni publican la información, sino que sólo la ponen a disposición; no se incluyen aquí las bibliotecas.

- **Administradores de sistemas:** responsables del manejo informático de los archivos, de los programas de recuperación de la información, de los equipos de computación e interconexión de redes, servicios de respaldo y recuperación, migración de datos, establecimiento de códigos de seguridad.
- **Bibliotecas:** aisladas o en consorcios, encargadas de poner los recursos de información a disposición de su comunidad de usuarios, se mantienen archivos digitales *in situ* o se conectan a proveedores comerciales.
- **Archivos digitales:** mantienen copias de los recursos electrónicos, una a la que se permite el acceso y otra copia de resguardo, distinguiendo entre ambos roles: uso y preservación.

Tipología

Según Ede (1999) existen dos tipos de recursos electrónicos:

- **en mano:** soportes físicos como disquetes, CD-ROM o cintas magnéticas;
- **en línea:** cualquier objeto que se puede encontrar en la Web: bases de datos, revistas electrónicas, páginas, listas de correo, grupos de noticia, sitios ftp, etc.

En cambio, Hedstrom (1998), los divide en:

- **creados originalmente en forma digital** y adquiridos a través de compra, donación o depósito legal; y
- **digitalizados** a partir de documentos impresos, o manuscritos convertidos a formato electrónico.

Una diferencia entre ambas categorías es que los primeros son más amorfos, menos controlados bibliográficamente.

Por su parte, Waters (1995), establece cinco categorías:

1. Origen y modo de distribución:

- de propiedad personal (e-mail, notas, manuscritos, pre-impresos, etc.);
- de propiedad corporativa (reportes de empresas, registros financieros, documentos de planificación, etc.);

- emanados de editoriales (libros, revistas, grabaciones, filmes);
- pertenecientes a bibliotecas, museos e instituciones educativas.

2. Contexto:

- fuentes secundarias (catálogos, índices, directorios, etc.);
- fuentes primarias (textos monográficos, imágenes, grabaciones, colecciones de datos).

3. Estructura de codificación, formato y contenido:

- codificados según formatos normalizados (ASCII, TIFF, etc.);
- con características de autodescripción (SGML con DTD);
- multimedia que incorporen varios formatos.

4. Cualidades dinámicas:

- pueden ser revisados y actualizados, de modo que existen versiones o ediciones;
- con cambios acumulativos o de manera interactiva, como por ejemplo las contribuciones a listas de discusión;
- variación de los enlaces entre un objeto y otros.

5. Uso:

- objetos que por falta de consulta son borrados, sin tener en consideración una posible demanda futura.

A su vez, Law (1996) los agrupa en cuatro categorías:

1. **Publicados comercialmente** (libros y revistas electrónicos, bases de datos, etc.). Este material es en general más fácil de describir.
2. **Documentos de trabajo** (literatura gris, pre-impresos, páginas web personales, etc.). Material difícil de controlar, no está asegurada su permanencia o acceso.

3. **Datos en bruto** (estadísticas, producidos por los gobiernos y organizaciones no gubernamentales, tomados de satélites). Muchas veces es información encriptada y puesta a disposición sólo de un grupo selecto de usuarios bajo ciertos términos.
4. **Archivos científicos** (trabajos y correspondencia entre investigadores), considerados muy importantes para el avance de una disciplina.