

Título: *Análisis y comparación de metadatos para la descripción de recursos electrónicos en línea*

Autores:

Silvina Marcela Angelozzi

Directora de Biblioteca
Universidad Blas Pascal
sangelozzi@ubp.edu.ar

Sandra Gisela Martín

Directora del Sistema de Bibliotecas
Universidad Católica de Córdoba
bibdir@uccor.edu.ar

RESUMEN

Este trabajo consiste en un estudio comparativo de metadatos para la descripción de documentos accesibles por Internet. Se analizan las particularidades y dificultades propias de los recursos electrónicos en línea en cuanto a su catalogación. Se definen y caracterizan los metadatos, se describen y analizan los diferentes esquemas. La comparación se realiza teniendo en cuenta características como origen y propósito, estructura y contenido de la descripción, complejidad, sintaxis, contribución a la normalización internacional, interoperabilidad, extensibilidad, flexibilidad, mantenimiento, documentación existente, actualización y resultados obtenidos hasta el presente.

PALABRAS CLAVE: CATALOGACION; INTERNET; METADATOS; RECURSOS ELECTRONICOS

INTRODUCCION

Actualmente cualquier catálogo de biblioteca puede nutrirse con 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 ofrecen 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 están fuera de la propia biblioteca, sí es una situación innovadora que comiencen a ser extraordinariamente accesibles. Más adelante, estos autores señalan la facilidad con 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. Así, 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 las posibilidades informativas.

Por ello, la integración de registros bibliográficos de materiales remotos se ha convertido en una creciente necesidad para la creación y mantenimiento de los catálogos de las bibliotecas.

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

Objetivo general: explorar los esquemas de metadatos que se perfilan como más apropiados para la catalogación de recursos electrónicos.

Objetivos específicos:

- analizar diferentes esquemas de metadatos para la descripción de recursos electrónicos accesibles por Internet;
- comparar los esquemas analizados en base a los criterios establecidos;
- analizar similitudes y diferencias entre la catalogación de materiales tradicionales y de recursos de información en Internet.

MARCO TEORICO

Las bibliotecas tienen una larga tradición en el desarrollo de sistemas de información, centrando su atención en organizar y proveer acceso a los documentos. Mediante el establecimiento de una normativa específica, han logrado poner en marcha 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 aseguran 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 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.

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

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 nuevos términos muy imprecisos como "materiales especiales", "nuevos documentos", "documentos no textuales". El problema se complica aún 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 fijas y en movimiento.

Actualmente cualquier documento se puede convertir a formato digital; así, cuando hablamos 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.

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.

El presente estudio se acota 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, de gobierno y de instituciones pioneras en redes digitales.

Así, los esquemas de metadatos seleccionados para la comparación son: MARC21/MARC XML, Dublin Core, MODS (*Metadata Object Description Schema*) y TEI (*Text Encoding Initiative*).

Se establece una grilla de comparación sobre la base de los criterios propuestos por Heery (1996), Dempsey & Heery (1998) y Taylor (2004):

- **Disciplina/ámbito de aplicación:** ¿quiénes están usando activamente este formato?. ¿Está asociado con una disciplina académica o interés profesional en particular?
- **Facilidad de creación de la descripción:** se refiere a la complejidad de creación del registro, si se requieren habilidades especiales, si están diseñados para que asigne los metadatos el autor, editor o 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:** en cuanto a 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:** referido a la posibilidad de incluir mayor nivel de detalle si es 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 cuál o cuáles son los que se perfilan con mayor posibilidad de éxito y difusión en el ámbito de la catalogación de recursos en red.

CATALOGACION DE RECURSOS ELECTRONICOS

Patrice 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 normas de la Descripción Bibliográfica Internacional Normalizada para recursos electrónicos, ISBD(ER) (1997), 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).

Las Reglas de Catalogación Angloamericanas, segunda edición, actualización 2003, AACR2R se ajustan a las ISBD (ER), adoptando la misma definición.

Un recurso electrónico de información puede ser un archivo, una base de datos, una página Web, un portal de Internet. Puede incluso existir por un lapso de tiempo, como la página de un evento, y luego desaparecer.

Robin Wendler (2000) presenta algunas características que dificultan su descripción:

- **Variabilidad en la presentación:** la apariencia y el orden del contenido no están normalizados y la presentación depende muchas veces del navegador.
- **Formas no familiares de publicación:** se asemejan a las publicaciones de hojas sueltas removibles, la nueva versión sustituye a la anterior y frecuentemente sin una numeración o indicación de que la versión ha cambiado.
- **Relaciones mal definidas de los materiales con otros soportes:** un subconjunto de las publicaciones electrónicas se deriva o está vinculado con documentos en otros soportes. No siempre es posible identificar el grado de coincidencia en cuanto al contenido o no existe consenso de cómo expresar estas relaciones.
- **Mutabilidad:** nada garantiza que un recurso en línea permanezca inalterable en el tiempo ni que su contenido no sea alterado sin previo aviso.
- **Movilidad:** no se puede asegurar la permanencia de un documento en el mismo sitio Web.
- **Ausencia de un receptor físico:** la falta de un usuario convencional elimina el punto clave sobre el cual la mayoría de las bibliotecas han construido los procedimientos de mantenimiento del catálogo.

A ello se puede agregar:

- **Gran variedad de recursos:** en Internet hay gran cantidad de textos, imágenes y sonidos, presentados en las más variadas formas.

Heery (1996) presenta además dificultades relacionadas con las siguientes características:

- **Ubicación:** un registro de un catálogo en línea de acceso público (OPAC) de una biblioteca se refiere a ubicaciones dentro de dicha institución. Los metadatos se

refieren a ubicaciones remotas; por lo tanto, se requerirán datos asociados al tipo de acceso (ftp, http) y a las restricciones de acceso (clave requerida). Muchas veces un recurso puede estar alojado en varios servidores (sitios espejo). En este sentido, los metadatos se asemejan a un registro de catálogo colectivo.

- **Versiones diferentes:** el mismo documento puede existir en diferentes formatos a la vez como Postscript, ASCII. Podría asemejarse a las distintas ediciones de un impreso.
- **"Granularity"**¹ : o nivel de detalle; otra particularidad de los recursos electrónicos en línea es cuando bajo un mismo título se agrupan múltiples recursos. Ejemplo de esto son las suscripciones a texto completo de un paquete de revistas; se plantea aquí la alternativa de catalogar individualmente cada título o simplemente el conjunto de ellos. Esto es simple cuando la cantidad de títulos es pequeña, pero se complica mucho para aquellas bases de datos de texto completo con miles y miles de revistas. El catalogador debe decidir a qué nivel analizar un documento: ¿se describe una base de datos o deben incluirse los registros individuales? En el caso de las páginas Web en varios niveles, por ejemplo, ¿cada vínculo se describe por separado?, ¿es un documento o varios?

Beall (1997) habla de **"volatilidad"** en que distingue tres tipos: **externa**, referida al cambio de URL o desaparición del sitio completo; **interna**, relacionada con el mantenimiento de los enlaces referidos en la página; y la **intelectual**, en cuanto a la actualización de contenidos.

Estas son algunas de las razones de por qué las nuevas formas de publicación alteran nuestros modelos establecidos para un control intelectual.

Los conceptos establecidos en las reglas de catalogación, tales como la fuente principal de información (portada), son difíciles de aplicar a estos documentos. Es necesario ampliar los modelos abstractos, las normativas y procedimientos de la bibliotecología, para poder describir las publicaciones electrónicas.

Por estas razones han surgido opiniones diversas entre los bibliotecólogos; así, hay quienes consideran que los recursos en línea son demasiado inestables como para ser efectivamente catalogados de la manera tradicional (Wendler, 2000; Méndez, 2002); otros esperan se desarrollen buscadores poderosos y herramientas automáticas de indización como para obviar la necesidad de catálogos; y algunos (Xu, 1996; Olson, 1997; Jul, 1997) llegaron a la conclusión que, a pesar de las numerosas dificultades que se encuentran al tratar de catalogarlos, los recursos de información disponibles en Internet deben ser catalogados como se viene haciendo con otros tipos de documentos para que puedan ser recuperados y utilizados eficientemente.

METADATOS

Existen diferentes definiciones sobre el término metadatos, siendo la más sencilla y de aceptación general: **"datos sobre los datos"**.

A continuación se presenta una serie de definiciones de diferentes autores:

- **Caplan (1995)** adopta la definición “datos sobre datos” y considera a las fichas de los catálogos tradicionales como metadatos ya que poseen datos bibliográficos: autor, título, editorial, etc., que se refieren a otros documentos.
- **Xu (1997)**, sostiene que es un conjunto de elementos que pueden ser usados para describir y representar objetos de información. En esta definición Xu habla de objetos de información, independientemente del medio o soporte.
- **Younger (1997)** entiende que los metadatos describen recursos, indican dónde están ubicados y qué se requiere para utilizarlos exitosamente.
- En el marco del **World Wide Web Consortium** (W3 Consortium), se los define como información sobre objetos de información en la Web, que puede ser leída por computadora (Swick, 1997).
- **Dempsey y Heery (1998)** los definen como datos que describen los atributos de un recurso.
- **Tennant (1998)**, define los metadatos como "información estructurada sobre información", y destaca que la palabra clave es "estructurada", ya que una descripción en texto libre no es suficiente, es necesario contar con ciertos elementos identificados formalmente y con una codificación para la especificación de una sintaxis dada.
- **Milstead y Feldman (1999)** consideran el registro bibliográfico en sí como metadatos y no necesariamente necesitarían estar codificados.
- Para **Wendler (2000)**, la definición es: " información necesaria para identificar, localizar, manejar y acceder a un recurso electrónico"
- **Gorman (2000)** diferencia lo que son estándares de estructura de aquellos que prescriben el contenido del registro bibliográfico. Para este autor, los metadatos han sido diseñados para responder a las necesidades de:
 - a) contar con una opción para la catalogación de los recursos electrónicos
 - b) encontrar una alternativa intermedia entre los altos costos de la catalogación "tradicional" y la simple búsqueda por palabra clave.

Para este autor no es necesario "inventar" algo nuevo ya que las técnicas de descripción bibliográfica de la catalogación tradicional cubren satisfactoriamente las necesidades de catalogación de los recursos electrónicos en línea, de acuerdo con las actualizaciones de las ISBD y de las Reglas de Catalogación Angloamericanas (AACR2R).
- **Taylor (2004)**, introduce dos conceptos, el de contenido y el de codificación, presentes ambos en la acepción del uso común del término metadatos. Cuando sólo hay contenido, se lo denomina un registro bibliográfico, cuando hay sólo codificación, es identificado como una "estructura". Para esta autora metadatos es una descripción codificada de un paquete de información.
- El documento ***Understanding metadata* (NISO, 2004)** define:

¹ La traducción de este término al español *granularidad* no resulta claro aplicado al contexto del tema que nos ocupa

Esquema de metadatos: es un conjunto de elementos de metadatos diseñados para un objetivo específico. El esquema define los nombres de los elementos y la semántica. Opcionalmente pueden determinar reglas de cómo asignar el contenido.

Semántica: constituye la definición o el significado cada uno de los elementos.

Contenido: es el valor asignado a los elementos de metadatos.

Codificación: representa la sintaxis o estructura. Los metadatos pueden ser codificados en diferentes sintaxis. Un esquema de metadatos puede tener reglas de sintaxis que indican cómo deben ser codificados sus elementos y puede no tener reglas en cuyo caso se denomina esquema con sintaxis independiente.

En este trabajo se entiende por metadatos, siguiendo a Taylor (2004), a la descripción codificada de un paquete de información como por ejemplo, un registro AACR2 codificado con MARC, un registro Dublin Core, etc. El propósito de los metadatos 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.

Sintetizando lo expuesto por Dempsey y Heery (1998), Lagoze (1996), Iannella y Waugh (1997), Taylor (2004), las funciones primarias de los metadatos son facilitar la identificación, ubicación, recuperación, manipulación y uso de recursos de información accesibles en línea.

TIPOLOGIA

De acuerdo con la naturaleza de los datos que describen, existen tres tipos de metadatos (Lazinger, 2001):

1. **descriptivos:** información bibliográfica
2. **estructurales:** referidos a formatos y estructuras
3. **administrativos:** incluye permisos, copyright y toda información relacionada al acceso.

Weibel (1997) distingue tres modelos para la creación y uso:

1. **modelo embebido** (*embedded model*), donde los metadatos son formulados por el creador del documento electrónico y la descripción reside incluida en el mismo. Esto es equivalente a la catalogación en la fuente o catalogación en publicación (CIP).
2. **generados por terceros** (*third-party metadata*), donde una agencia crea, reúne y administra los registros, pero no están incluidos en el recurso en sí. Se generarían productos similares a las bases de datos referenciales. Al registro, si bien está separado del recurso, se le puede agregar un vínculo para acceder directamente a él.
3. **visualización filtrada** (*view-filter mode*), una agencia administra registros de metadatos, con diferentes estructuras y los traduce en una descripción común compartida por todos los usuarios. Tiene la gran ventaja de proveer una interface única de búsqueda.

De acuerdo con la riqueza de la descripción y nivel de detalle, se los divide en tres niveles (Dempsey & Heery, 1997):

Nivel 1: datos no estructurados, extraídos automáticamente de los documentos e indizados para la búsqueda.

Nivel 2: estructurados, son relativamente nuevos y muchos han sido diseñados ad hoc para su uso en Internet. En principio no estuvieron normalizados, pero actualmente algunos de ellos están surgiendo como normas, soportan búsquedas por campo y son relativamente fáciles de aplicar, están pensados para que los incluya el autor del documento. Ej. Dublin Core

Nivel 3: ricos para la descripción de documentos, diseñados para ser aplicados por profesionales de las ciencias de la información, son complejos y basados en normas internacionales. Su nivel de detalle implica alta prestación en la recuperación y dificultad en su aplicación. Ej. formato MARC.

ANÁLISIS DE LOS ESQUEMAS DE METADATOS

Como se citó anteriormente, los esquemas de metadatos que se comparan en este trabajo son MARC21/MARC XML, Dublin Core, MODS y TEI.

Para elaborar la grilla de comparación se adoptaron los siguientes criterios que surgen de la adaptación de los ya propuestos por Heery (1996), Dempsey & Heery (1998) y Taylor (2004).

1. **Año de Origen**
2. **Organismo creador**
3. **Mantenimiento:** se refiere especialmente al organismo encargado del desarrollo, actualización y mantenimiento del esquema de metadatos.
4. **Documentación existente:** existencia de manuales, tutoriales, listas de distribución, etc.
5. **Objetivos/misión:** la finalidad con que han sido creados.
6. **Disciplina/ámbito de aplicación:** área académica o grupo profesional que utiliza el formato de metadatos.
7. **Corriente teórica a la que responden:** si proviene de la comunidad bibliotecaria, informática, de editores o proveedores.
8. **Simplicidad/complejidad:** en cuanto a la facilidad de creación del registro, si se requieren habilidades especiales, si están diseñados para que asigne los metadatos el autor, editor o proveedores centralizados.
9. **Interoperabilidad:** entendida según la definición del documento *Understanding metadata*. (NISO, 2004) como la habilidad de sistemas múltiples con diferentes plataformas de hardware y software, estructuras de datos e interfases, para intercambiar datos con una mínima pérdida de contenido y funcionalidad.
10. **Extensibilidad:** posibilidad de incluir mayor nivel de detalle si es necesario mediante campos opcionales, calificadores, etc.
11. **Flexibilidad:** posibilidad de elegir los campos/elementos a incluir.

12. **Estructura:** elementos o campos de la codificación y características de los mismos.
 - a. **Opción:** si contiene campos/elementos que son opcionales y obligatorios.
 - b. **Repetición:** si admite campos y subcampos repetibles.
 - c. **Subcampos:** admisión o no de subcampos.
 - d. **Control:** posibilidad de controlar valores en algunos campos.
13. **Contenido:** si prescriben cómo debe ser elaborada la descripción en cada campo o si se valen de normativas externas, como por ej. AACR2.
14. **Nivel de catalogación permitida:** se refiere al nivel de detalle de catalogación permitido, asociado a nivel de complejidad 1, 2 o 3, que va desde los formatos simples, a los ricos o muy estructurados.
15. **Actualización:** últimas versiones o modificaciones disponibles.
16. **Contribución a la normalización internacional:** protocolos de Internet asociados y clases de normas que incluyen, y si en sí mismos se han consolidado como normas de hecho o de derecho. Por normas de hecho se entiende a las adoptadas por organismos o asociaciones profesionales (IFLA, ALA, etc.) y por normas de derecho a las adoptadas por organismos de normalización (ISO, AENOR, etc.)
17. **Sintaxis:** se evalúa aquí si el esquema de metadatos aplica una sintaxis que indica cómo deben ser codificados sus elementos.
18. **Resultados obtenidos hasta el presente:** se refiere al grado de difusión del esquema de metadatos en cuanto a aplicaciones concretas.

A continuación se analiza cada esquema de metadatos, según las categorías descriptas.

MARC21/MARCXML

Año de Origen: década de 1960. En 1992 se incluyó el campo 856 y la versión inicial de MARC XML, apareció en 2002.

Organismo creador: Library of Congress

Mantenimiento: a cargo de la Network Development and MARC Standards Office, Library of Congress.

Documentación existente: manuales en línea en varios idiomas, tutoriales y listas de distribución.

Objetivos/misión: ha sido creado inicialmente para la catalogación de documentos impresos, específicamente obras monográficas, y luego fue incorporando campos para otro tipo de recursos: publicaciones seriadas, archivos de computadora, mapas, música, registros sonoros, manuscritos, material audiovisual y recursos electrónicos.

Disciplina/ámbito de aplicación: se utiliza para describir documentos en distintos formatos y para todas las disciplinas, empleado especialmente por la comunidad bibliotecaria.

Corriente teórica a la que responde: proviene de la comunidad bibliotecaria; por lo tanto, es ampliamente utilizado en ese ámbito. Se puede decir que proviene de una corriente estructuralista.

Simplicidad/complejidad: MARC es un formato complejo; para la creación de un registro bibliográfico completo se requiere personal especializado y capacitado, con conocimiento profundo de las reglas de catalogación. Alto costo de la descripción dada su complejidad.

Interoperabilidad: puede ser exportado a distintos formatos bibliográficos y de metadatos.

Extensibilidad: permite incluir mayor nivel de detalle si es necesario mediante campos opcionales, o el empleo de ciertos subcampos.

Flexibilidad: MARC tiene poca flexibilidad a la hora de elegir los campos para una descripción bibliográfica. Hay campos obligatorios establecidos.

Estructura:

El formato MARC posee más de 900 campos. En los últimos años se ha actualizado para incluir características propias de los recursos electrónicos con los siguientes campos:

538 requerimientos del sistema y modo de acceso - Nota

256 características del archivo digital

516 tipo de archivo o dato digital - Nota

856 Acceso y localización electrónica

Los tres primeros campos (256, 538, 516) se aplican tanto a archivos digitales locales como remotos. El campo 856, Acceso y localización electrónica, fue desarrollado especialmente para los recursos electrónicos disponibles en la Web o a través de redes. Utiliza etiquetas numéricas.

- a. Opción: contiene campos que son opcionales y obligatorios
- b. Repetición: en teoría todos los campos y subcampos son repetibles, pero la naturaleza de ciertos datos hace que determinados campos o subcampos sean no repetibles como por ejemplo el campo 100 (campo no repetible) \$a (subcampo no repetible) \$c (subcampo repetible).
- c. Subcampos: contiene numerosos subcampos
- d. Control: recomienda controlar valores en algunos campos, por ejemplo en el uso de vocabularios controlados y en el control de autoridades.

Contenido: MARC es en sí un estándar de estructura. El contenido de la mayoría de los elementos está definido por normativas fuera del formato en sí mismo, por ejemplo: *Anglo-American Cataloguing Rules*, *Library of Congress Subject Headings*, *National Library of Medicine Classification*.

Nivel de catalogación permitida: muy completa. Se ubica en Nivel 3 (Dempsey, 1997). Están previstos dos niveles definidos para los Estados Unidos: un nivel mínimo y el nivel completo o nacional que contiene la suficiente información catalográfica para ser utilizada por agencias de catalogación tanto nacionales como internacionales.

Actualización: la última actualización corresponde al año 2006.

Contribución a la normalización internacional: es estándar de hecho para la representación de registros bibliográficos. La estructura de los registros MARC son una implementación de normativas nacionales e internacionales: *Information Interchange Format* (ANSI Z39.2) y *Format for Information Exchange* (ISO 2709) respectivamente.

MARC utiliza además, los siguientes estándares:

- ISO 2108: International Standard Book Numbering (ISBN) (ANSI/NISO Z39.21)
- ISO 3901: International Standard Recording Code (ISRC)
- ISO 3297: International Standard Serial Number (ISSN) (ANSI/NISO Z39.9)
- ISO 10324: Holdings Statements - Summary Level (ANSI/NISO Z39.71)

Sintaxis: se puede codificar en distintas sintaxis: HTML, XML.

Resultados obtenidos hasta el presente: en cuanto a recursos electrónicos es utilizado por las bibliotecas que participan de los diferentes proyectos de OCLC (InterCat, CORC, actualmente absorbidos por Connexion), y que incluyen recursos Web en sus OPACs.

DUBLIN CORE

Año de Origen: 1995

Organismo creador: Online Computer Library Center (OCLC) y National Center for Supercomputing Applications (NCSA)

Mantenimiento: The Dublin Core Metadata Initiative (DCMI)

Documentación existente: manuales en línea.

Objetivos/misión: Dublin Core (DC) ha sido creado ad-hoc para la catalogación de documentos electrónicos con el objetivo de definir un conjunto de elementos básicos para que los propios autores puedan describir sus recursos Web, de modo que puedan ser recuperados por los motores de búsqueda de Internet. Entre los propósitos también están definir el marco para la interoperabilidad entre conjuntos de metadatos y apoyar el desarrollo de elementos específicos para ciertas disciplinas.

Disciplina/ámbito de aplicación: se utiliza para describir documentos en cualquier ámbito o materia. DC es aplicado por diferentes comunidades fuera de la bibliotecología tales como educación, gobierno, archivos, museos.

Corriente teórica a la que responden: proviene de la interacción de la comunidad bibliotecaria e informática.

Simplicidad/complejidad: en su origen es simple; son sólo 15 elementos y está pensado para que el autor pueda incluir los metadatos al momento de crear el documento. Existen dos corrientes: los *minimalistas*, que bogan por mantener el nivel de simplicidad y los *estructuralistas*, que proponen un mayor nivel de distinción y extensibilidad. De allí que surjan las posturas de utilizar DC con calificadores o sin calificadores. No requiere de catalogadores expertos. Alta productividad debido a su sencillez.

Interoperabilidad: puede ser exportado a distintos formatos bibliográficos y de metadatos.

Extensibilidad: permite incluir mayor nivel de detalle si es necesario mediante los calificadores.

Flexibilidad: nada en el DC es obligatorio, todos los elementos son opcionales; así el usuario elige la profundidad de una descripción.

Estructura: contiene 15 elementos básicos agrupados en tres grandes grupos: contenido, propiedad intelectual e instalación/manipulación. Utiliza etiquetas basada en palabras y no numéricas.

- a) Opción: todos los elementos son opcionales
- b) Repetición: todos los elementos son repetibles.
- c) Sucampos: no posee subcampos pero sí calificadores.
- d) Control: no requiere control de valores en los elementos.

Contenido: es un estándar de estructura. No está regido por reglas de contenido, sino que se sugiere la normativa a aplicar, por ejemplo para el elemento Idioma, se sugiere utilizar los códigos de la ISO 639-2.

Nivel de catalogación permitida: sencilla, concisa. Se lo ubica en un Nivel 2 (Dempsey, 1997).

Actualización: la última actualización corresponde al año 2007.

Contribución a la normalización internacional: es norma de derecho, desde octubre de 2001 (ANSI/NISO Standard Z39.85-2001). Las versiones actualizadas son:

- ISO Standard 15836-2003 (Febrero 2003): <http://www.niso.org/international/SC4/n515.pdf>
- ANSI/NISO Standard Z39.85-2007 (Mayo 2007): The Dublin Core Metadata Element Set <http://www.niso.org/standards/resources/Z39-85-2007.pdf>

Sintaxis: posee independencia sintáctica. DC se puede codificar en distintas sintaxis como HTML, XML o RDF.

Resultados obtenidos hasta el presente: son muchos los proyectos en marcha basados en DC en diversas áreas del conocimiento. Se han desarrollado numerosas aplicaciones informáticas para la asignación de metadatos en DC. Es el estándar más divulgado.

MODS (*Metadata Object Description Schema*)

Año de Origen: 2002

Organismo creador: Library of Congress

Mantenimiento: Network Development and MARC Standards Office, Library of Congress.

Documentación existente: manuales en línea y lista de distribución.

Objetivos/misión: es un derivado de MARC21 bibliográfico, creado para permitir la extracción de datos de manera selectiva de los registros MARC21 existentes y crear registros originales.

Disciplina: utilizado para distintos propósitos pero fundamentalmente para aplicaciones bibliotecarias.

Corriente teórica a la que responden: proviene de la comunidad bibliotecaria.

Simplicidad/complejidad: se puede decir que MODS posee una complejidad intermedia. Es más complejo que DC y está más orientado a bibliotecas. Pero es más simple que MARC.

Interoperabilidad: puede ser exportado a distintos formatos bibliográficos y de metadatos al igual que MARC. Mapeos disponibles: MARCXML a MODS ; MODS a MARCXML; Dublin Core a MODS ;MODS a Dublin Core ;MODS Version 2.0 a MODS Version 3.0; MODS 3.0 a HTML.

Extensibilidad: permite extender la descripción conectando con descripciones alojadas fuera del registro, por ej. un abstract disponible en otra dirección Web.

Flexibilidad: la flexibilidad está dada al ser opcionales todos sus elementos.

Estructura: incluye un subconjunto de campos de MARC21 bibliográfico. La mayoría de los elementos de MODS tiene su equivalente en MARC21 aunque también tiene incorporado otros campos específicos para la descripción de recursos digitales. Consta de 20 elementos principales, subelementos y atributos que pueden ser usados en uno o más elementos.

Utiliza etiquetas basadas en palabras y no numéricas.

- a) Opción: todos los elementos y atributos son opcionales.
- b) Repetición: los elementos son repetibles y los atributos no repetibles.
- c) Subcampos: contiene subelementos y atributos.
- d) Control: en ciertos elementos el atributo **“authority”** permite registrar con qué herramienta se realiza el control del valor.

Contenido: es un estándar de estructura, el contenido de la mayoría de los elementos se anota de acuerdo a normativas externas (por ej. AACR2) de manera análoga a MARC21.

Nivel de catalogación permitida: más sencillo que MARC21 y más orientado al usuario que MARC XML. Permite la conversión de los campos centrales de un registro MARC21, aunque pueden perderse algunos datos específicos. En el caso de una descripción original, permite crear un registro simple, en algunos casos usando mayor cantidad de etiquetas generales que las disponibles en un registro MARC. Atiende a la granularidad, se puede describir el todo o parte de un recurso.

Actualización: la última versión es MODS 3.2 (2006). Actualmente está en revisión la versión 3.3.

Contribución a la normalización internacional: en 2004 ha sido registrado en NISO. Un punto importante a la hora de convertirse en estándar es la total compatibilidad con MARC21.

Sintaxis: se codifica en XML.

Resultados obtenidos hasta el presente: en la Web de la Library of Congress, hay 25 proyectos listados de implementaciones de este esquema. Entre las instituciones figuran la LC, la National Library of Australia, la Digital Library Federation, varias universidades de Estados Unidos, Oxford University.

TEI (*Text Encoding Initiative*)

Año de Origen: el proyecto se inicia en 1987; en 1994 se publican las Recomendaciones TEI (*TEI Guidelines*) y en diciembre del 2000 se crea un consorcio para su mantenimiento (*TEI-C*).

Organismo creador: el proyecto TEI surge de la investigación cooperativa entre tres asociaciones: Association of Computers in the Humanities (ACH), Association for Computational Linguistics (ACL), y Association of Literary and Linguistic Computing (ALLC) y con fondos de: U.S. National Endowment for the Humanities (NEH), European Community, Mellon Foundation y Social Science and Humanities Research Council of Canada.

Mantenimiento: en el año 2000 se establece una nueva corporación sin fines de lucro denominada TEI Consortium, cuya oficina ejecutiva se ubica en Charlottesville, Virginia y con sedes en Brown University, Oxford University y University of Virginia.

Documentación existente: manuales y tutoriales en línea, software, varios grupos de discusión, presentaciones para uso educativo y un wiki.

Objetivos/misión: TEI ha sido desarrollado como un estándar internacional e interdisciplinario para bibliotecas, museos, editoriales corporativas y editores individuales para representar una gran variedad de textos literarios y lingüísticos para la investigación, preservación, y búsqueda en línea. Se trata de representar los materiales textuales en formato electrónico de tal forma que permita a los investigadores de cualquier disciplina intercambiar y reutilizar los recursos independientemente del hardware y software que utilicen.

Disciplina/ámbito de aplicación: se aplica especialmente en el campo de la lengua, la literatura, las humanidades y las ciencias sociales. Se utiliza para describir textos literarios y lingüísticos en bibliotecas, museos y editoriales. Sin embargo, puede aplicarse para la estructuración de información electrónica de todo tipo, incluso imágenes.

Corriente teórica a la que responden: proviene de la comunidad de estudio de la lingüística.

Simplicidad/complejidad: la complejidad está asociada a la cantidad de información que se quiera incluir y el nivel de normalización que se pretenda. Cada proyecto que utilice TEI debe tomar sus propias decisiones respecto a qué partes del esquema utilizar y el nivel de complejidad o sofisticación a emplear. Requiere del conocimiento del modelo y de las reglas bibliotecarias.

Interoperabilidad: ha sido concebido para intercambio, con independencia de software y hardware.

Extensibilidad: permite un marcado semántico muy detallado.

Flexibilidad: en los instructivos se sugiere que el usuario adecue TEI a sus necesidades, para obtener la máxima prestación.

Estructura: se partió en 1994 con la definición de 600 elementos en SGML, luego surgieron las versiones TEI P3 (1998) más operativa y modular, TEI P4 (2001) en XML y TEI Lite orientada a bibliotecas digitales. Es más completo que DC y está más orientado a bibliotecas. Los elementos están estructurados en 4 secciones: la **descripción de archivo**, que contiene la descripción bibliográfica del texto similar a AACR2R e ISBD; **descripción de codificación**, con elementos sobre características del archivo electrónico, **descripción de perfil**, con información similar a los puntos de acceso secundarios de AACR2R, datos de lengua, clasificación, materias y **descripción de revisión** que contiene un registro de cada cambio que se ha hecho al texto y la fecha en que ha sido cambiado. Utiliza etiquetas basada en palabras y no numéricas.

- a) Opción: existen campos obligatorios, opcionales y recomendados.
- b) Repetición: posee elementos repetibles y no repetibles.
- c) Subcampos: no se utilizan subcampos, sí atributos dentro de los campos.
- d) Control: se sugiere establecer control para ciertos elementos (Ej. Series).

Contenido: es un estándar de estructura que en ciertos elementos prescribe el contenido, está implícito el uso de AACR2 O ISBD. Permite además información en texto libre.

Nivel de catalogación permitida: formato rico, se ubica en un Nivel 3 (Dempsey, 1997).

Actualización: la versión actual de esquema es TEI P4 del año 2004. Actualmente se está preparando la versión TEI P5.

Contribución a la normalización internacional: se puede considerar que actualmente es un estándar de hecho.

TEI hace referencia a los siguientes estándares:

- ISO 646: Information technology -- ISO 7-bit coded character set for information interchange
- ISO 10646: Information technology -- Universal Multiple-Octet Coded Character Set (UCS)
- Unicode

Sintaxis: se codifica en SGML y XML.

Resultados obtenidos hasta el presente: actualmente existente 138 proyectos registrados en la página Web de TEI.

COMPARACION DE LOS ESQUEMAS

En cuanto a su **origen** MARC, MODS y Dublin Core surgen de instituciones ligadas a la bibliotecología como son LC y OCLC. TEI proviene del ámbito de la lingüística, pero para diseñar la descripción de archivo se interactuó con bibliotecarios, entre otros profesionales.

MARC es el formato más antiguo (1960), TEI se inicia en 1987 y en 1994 se publican las primeras recomendaciones, posteriormente aparece Dublin Core (1995) y finalmente MODS nace en 2002. Se puede deducir que se ha partido de formatos sumamente complejos como MARC y TEI para luego pasar a un formato más simple como DC y finalmente surge un formato intermedio como lo es MODS. Todos provienen de instituciones altamente reconocidas, está previsto su mantenimiento y desarrollo y cuentan con documentación detallada en línea.

En cuanto al **propósito**, MARC ha sido creado para catalogación de documentos impresos y se fue extendiendo para los nuevos soportes, para el caso de recursos electrónicos se incorporaron nuevos campos, pero siempre respondiendo al mismo modelo. El último avance ha sido la codificación en XML para adaptarlo al ambiente Web. Dublin Core, TEI y MODS, han sido creados ad-hoc para descripción de recursos electrónicos. Dublin Core y TEI priorizando la facilidad de recuperación a través de motores de búsqueda y el intercambio de información. MODS apunta a una descripción más simple y flexible que MARC21, pero a la vez totalmente compatible con él.

La **interoperabilidad** es un criterio de fuerte presencia en todos los formatos y pueden codificarse en **sintaxis** XML, estándar no propietario.

En cuanto a la **complejidad** de creación de los registros, DC es sin duda el más sencillo, pensado para que el propio autor pueda asignar los metadatos en el momento de creación del archivo digital. De todas maneras, se complejiza, al aumentar el nivel de detalle de la descripción a través del empleo de los calificadores. MARC y TEI son complejos por el nivel de detalle, cantidad y alta estructuración de los elementos, y están asociados a las normativas bibliotecarias para registro del contenido de la descripción. MARC de manera obligatoria y TEI, si bien deja opción, sugiere la aplicación de las AACR2 e ISBD en varios de sus elementos. MODS estaría en un intermedio, al tener menos elementos, pero se requiere conocimiento de las normativas bibliotecarias, porque los elementos se equiparan a MARC.

La simplicidad relativa, la flexibilidad y la compatibilidad con MARC, dan a MODS muchas posibilidades para las bibliotecas.

A excepción de MARC, todos son **flexibles** y **extensibles** a la hora de seleccionar los elementos a incluir, el usuario decide el **nivel de detalle de la descripción**. En cuanto a esta última característica MARC y TEI, según la clasificación de Dempsey & Heery (1997), se ubican en un nivel 3, son formatos ricos altamente estructurados. Dublin Core se ubica en la banda 2, con una descripción sencilla. MODS estaría moviéndose entre los niveles 2 y 3 dependiendo si se incluyen o no todos los subelementos y atributos.

A MODO DE CONCLUSION

A nuestro juicio los diferentes esquemas de metadatos van a coexistir, se aplicarán unos u otros de acuerdo con las necesidades, por lo que el desarrollo deberá ir por el camino de la interoperabilidad, para posibilitar el intercambio entre los diferentes esquemas. En este sentido, *Resource Description Framework* (RDF), si bien no se desarrolla en este trabajo, es un marco adecuado para lograr la interoperabilidad de los diversos modelos de metadatos, en el ambiente Web basado en XML.

Se pasaría del catálogo único y centralizado a uno distribuido, con metadatos asociados a los documentos, registros adicionales creados por bibliotecas y centros de información, que puedan ser recuperados e interpretados por motores de búsqueda.

Creemos muy potente la opción de los metadatos embebidos, que nacen con el documento electrónico, como catalogación en fuente o “portada electrónica normalizada” a través de una cabecera. De todos modos, no será tarea fácil conseguir que autores y editores agreguen esta información y lo hagan de manera normalizada, pero un documento en la red sin metadatos, tendrá escasa o nula visibilidad. Tendrá tal vez que surgir software de asistencia para asignación de metadatos de manera ágil.

La catalogación de recursos electrónicos disponibles en la Web requiere un trabajo interdisciplinario entre los especialistas autores del contenido de los documentos como biólogos, matemáticos, ingenieros, en estrecha vinculación con informáticos y bibliotecarios.

BIBLIOGRAFIA CONSULTADA

- Beall, J.** (1997). Cataloging world wide Web sites consisting mainly of links. *Journal of Internet Cataloging*, 1 (1), 83-92
- Burnett, K.,** Bor Ng, K. & Park, S. (1999). A comparison of the two traditions of metadata development. *JASIS*, 50 (13), 1209-1217
- Caplan, P.** (1997). To Hel(sinki) and back for the Dublin Core. *Public Access Computer Systems Review*. 8 (4). Recuperado el 10 de octubre de 2007, de <http://info.lib.uh.edu/pr/v8/n4/capl8n4.html>
- Caplan, P.** (1995). You call it corn, we call it syntax-independent metadata for document-like objects. *The public access computer systems review*, 6 (4), 19-23
- Dempsey, L.** & Heery, R. (1998). Metadata : a current view of practice and issues. *Journal of Documentation*, 54 (2), 145-172
- Dovey, M. J.** (1999). "Stuff" about "Stuff" the different meanings of metadata. *Vine*, 116, 6-13
- Dublin Core Metadata Initiative** (2007). Recuperado el 22 de octubre del sitio Web oficial del DCMI, <http://dublincore.org/>
- Gartner, R.** (2003). MODS: Metadata Object Description Schema. Recuperado el 28 de octubre de 2007, de http://www.jisc.ac.uk/uploaded_documents/tsw_03-06.pdf
- Gimeno Montoro, M.J.,** Barrueco Cruz, J.M. & Garcia Testal, C. *Catalogación de recursos electrónicos accesibles en Internet: revisión de propuestas para una normativa*. FESABIB 98. VI Jornadas Españolas de Documentación. Recuperado el 23 de octubre de 2000, de http://www.florida-uni.es/~fesabid98/Comunicaciones/mj_gimeno.htm
- Gorman, M.** (2000). *¿Metadatos o catalogación? Un cuestionario erróneo*. En: F.F.Martínez Arellano & L. Escalona Ríos (Comps.). *Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica* (pp.1-20). México : Universidad Nacional Autónoma.
- Heery, R.** (1996). Review of Metadata Formats. *Program*, 30(4), 345-373.
- Heery, R.** (1998). What is ... RDF? . *Ariadne*, 14. Recuperado el 10 de octubre de 2007, de <http://www.ariadne.ac.uk/issue14/what-is/>
- Hopkins, J.** (2000). *USMARC como estructura de metadatos*. En: F.F. Martínez Arellano & L. Escalona Ríos, L. (Comps.). *Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica* (pp.21-35). México : Universidad Nacional Autónoma.
- Hsieh-Yee, I.** (2000). Organizing Internet resources: teaching cataloging standards and beyond. *OCLC Systems & Services*, 16 (3), 130-145
- Iannella, Renato** & Waugh, Andrew.(1997). *Metadata. Enabling the Internet* . Recuperado el 23 de enero de 2007, del sitio Web de IFLA: <http://www.ifla.org/documents/libraries/cataloging/metadata/ianr1.pdf>
- IFLA.** Universal Bibliographic Control and International MARC Programme. (1997) *International standard bibliographic description for electronic resources: ISBD(ER)*; revised from the ISBD(CF): International standard bibliographic description for computer files. München : K.G.Saur.
- Jul, E.** (1997). Cataloguing Internet resources: surveys and prospectus. *Bulletin of the American Society for Information Science*, 24 (1), 6-9
- Lazinger, S.** (2001). *Digital preservation and metadata: history, theory, practice* Englewood, Colorado: Libraries Unlimited
- Lyons, P.** (1997). Managing access to digital information: some basic terminology issues. *International Information & Library Review*, 29, 207-208
- MARC Standards.** (2007). Recuperado el 15 de octubre de 2007, del sitio Web de la Library of Congress <http://www.loc.gov/marc/>
- Martínez Arellano, F.F.** & Escalona Ríos, L. (Comps.) (2000). *Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica*. México : Universidad Nacional Autónoma.
- Méndez Rodríguez, E.** (2002). *Metadatos y recuperación de información*. Gijón : Trea.
- Milstead, J.** & Feldman, S. (1999). Metadata: cataloging by any other name. *Online*, 23 (1), 24-31
- MODS.** Metadata Object Description Schema official Web site (2007). Recuperado el 29 de octubre de 2007, del sitio Web de la Library of Congress <http://www.loc.gov/standards/mods/>
- NISO Standard Z39.85-2007** (2007): The Dublin Core Metadata Element Set <http://www.niso.org/standards/resources/Z39-85-2007.pdf>
- Olson, N.B.** (1997) *Cataloging Internet resources: a manual and practical guide*. 2nd. Ed. Recuperado el 15 de febrero de 2002, del sitio Web de OCLC <http://www.oclc.org/oclc/man/9256cat/toc.htm>
- Pinto Molina, M.,** García Marco, F.J. & Agustín Lacruz, M. (2002). *Indización y resumen de documentos digitales y multimedia*. Gijón : Trea.

- Real Academia Española.** (1991). *Diccionario de la lengua española*. (20ª. ed.). Madrid : Real Academia Española.
- Reglas de catalogación angloamericanas** (2ª. Ed, rev. 2002, act. 2003) **(2004)**, preparadas bajo la dirección de Joint Steering Committee for Revision of AACR.. Bogotá: Rojas Eberhard
- Sha, V.T.** (1995). Cataloguing Internet resources: the library approach. *The Electronic Library*, 13 (5), 467-476
- Sperberg-McQueen, C.M.** & Burnard, L. (Eds.) (2004). *TEI P4: Guidelines for Electronic Text Encoding and Interchange XML-compatible edition*. The TEI Consortium. Recuperado el 31 de octubre de 2007 del sitio Web oficial de TEI <http://www.tei-c.org/cms/Guidelines/P4/html/index.html>
- Swick, R.** (1997). *Metadata: A W3C Activity*. Recuperado el 24 de enero de 2007 del sitio Web del W3 Consortium <http://www.w3.org/Metadata/Activity.html>.
- Taylor, A.** (2004). *The Organization of Information*. (2ª ed.) Englewood, CO: Libraries Unlimited
- TEI** The Text Encoding Initiative (2007). Recuperado el 25 de octubre de 2007, del sitio Web del TEI Consortium <http://www.tei-c.org/>
- Tennant, R.** (1998). 21st Century cataloguing. *Library Journal*, 123 (7), 30-31
- Vellucci, S.** (1998). Metadata. *ARIST*, 33, 187-222
- Weibel, S.** (1997). The Dublin Core: a simple content description model for electronic resources. *Asis Bulletin*, 24 (1). Recuperado el 10 de octubre de 2007, de <http://www.asis.org/Bulletin/Oct-97/weibel.htm>
- Weibel, S.** & Dekkers, M. (2002). Dublin Core Metadata Initiative Progress Report and Workplan for 2002. *D-Lib Magazine*, 8 (2). Recuperado el 10 de octubre de 2007, de <http://www.dlib.org/dlib/february02/weibel/02weibel.html#NISO-Z39.85>
- Wendler, R.** (2000). *Diversificación de actividades: habilidades y funciones catalográficas en la era digital*. En: F.F. Martínez Arellano & L. Escalona Ríos (Comps.). *Internet, metadatos y acceso a la información en bibliotecas y redes en la era electrónica* (pp.36-48). México: Universidad Nacional Autónoma.
- Woodward, J.** (1996) Cataloging and classifying information resources on the Internet. *ARIST*, 31 , 189-220
- Xu, Amanda.** (1997). Metadata conversion and the library OPAC. Recuperado el 12 de octubre de 2007, del sitio Web de IFLA: <http://www.ifla.org/documents/libraries/cataloging/metadata/xu.pdf>
- Younger, J.** (1997). Resources description in the digital age. *Library trends* 45, 462-81