

PRINCIPALES SISTEMAS DE METADATOS PARA LA DESCRIPCIÓN BIBLIOGRÁFICA. CRITERIOS PARA SU DISEÑO, EVALUACIÓN Y MANTENIMIENTO

Marina Arana Montes. Servicio de Proceso Técnico y Normalización. BUC

METADATOS

DEFINICIONES

El término se ha introducido intuitivamente, sin una definición formal. Por ello hoy existen varias definiciones. La más común es la traducción literal:

Un metadato es un dato acerca de otro dato, ya que proporciona la información mínima necesaria para identificar un recurso; es decir, "información más específica (dato) acerca de un contenido particular (dato). Puede incluir información descriptiva sobre el contexto, calidad y condición o características del dato.

Como para la mayoría de las personas la diferencia entre dato e información es algo meramente filosófico e irrelevante a efectos prácticos, otras definiciones pueden ser:

- Un metadato es información sobre datos
- Un metadato es información sobre información (Información estructurada sobre la información que se ofrece o distribuye)

Hay definiciones más sofisticadas, por ejemplo:

- "Un metadato es un dato estructurado y codificado, que describe características de entidades portadoras de información (datos) para ayudar a la identificación, búsqueda, evaluación y gestión de dichas entidades"
- "[Un metadato es una serie de] descripciones opcionales estructuradas que están disponibles públicamente para ayudar explícitamente a localizar objetos
- "Un metadato es toda aquella información descriptiva sobre el contexto, calidad, condición o características de un recurso, dato u objeto, que tiene la finalidad de facilitar su recuperación, autenticación, evaluación, preservación y/o interoperabilidad (SEDIC)

Éstas definiciones tienden a concentrarse en un sólo propósito de los metadatos -encontrar "objetos", "entidades", o "recursos"- desentendiéndose de otros, como el uso de metadatos para optimizar algoritmos de compresión o ejecutar cálculos adicionales usando el dato.

Cualquier tipo de dato es la descripción de algo. El metadato es un tipo de dato en el que lo descrito es otro dato.

Ejemplos:

<i>Dato</i>	<i>Metadato</i>
30°	Grados (centígrados/fahrenheit) (no significa lo mismo si es fiebre o del aire)
2503467J	DNI
2.782	Saldo (euros/dólares/libras)
6,5	Nota media
Rebajado	Precio

A su vez, el metadato, como dato que también es, puede tener su propio metadato(s). Estos suelen llamarse no muy originalmente meta-metadatos.

Ejemplo:

Dato: 28009

Metadato: Código Postal

Dato: 27-09-07

Metadato: actualización

La jerarquía dato, metadato, meta-metadato, puede continuar indefinidamente, desde un punto de vista teórico. Afortunadamente, tenemos los suficientes conocimientos acumulados como para comprender y utilizar cualquier dato, definido formalmente, con unos pocos metadatos.

CARACTERÍSTICAS

El término no ha evolucionado mucho desde su creación en los 60, pero si lo han hecho enormemente los conceptos asociados, sus aplicaciones y nivel de uso.

Metadato es un término neutral, sujeto a la multidisciplinariedad de su creación por parte de diferentes profesionales y de su aplicación a diversos ámbitos. Ahora bien, en el ámbito que nos ocupa, que es el relacionado con las bibliotecas y los sistemas de información documental, la mayor parte de las tareas profesionales que se realizan (la catalogación entre ellas, aunque no solo) pueden considerarse sin duda, como procesos de asignación de metadatos.

En el contexto de una biblioteca, donde los datos son el contenido de los títulos almacenados, los metadatos acerca de una obra deberían incluir necesariamente una descripción del contenido, el autor, la fecha de publicación y la ubicación física, entre otros. En el contexto de la Web, los documentos digitales con metadatos se recuperan mejor que con los motores de búsqueda a texto completo.

Los metadatos específicos para los DOL (documentos digitales) en la web tendrían que incluir su propia identificación como tales (URL, DOI), su formato (html, tiff, jpg), sus condiciones de acceso y uso, etc.

NIVELES

Como se ha indicado, hay jerarquías de datos y metadatos. Sin embargo, cualquier dato particular puede situarse en distintos niveles de jerarquías según el contexto o el campo que se esté considerando. Por ejemplo, al tratar de la geografía de Madrid, "28009" sería un dato y "Código Postal" el metadato. Pero al estudiar la gestión de datos de un sistema automático que trata información geográfica, "Código Postal" puede ser el dato y "nombre del dato" y "5 caracteres de 0 a 9" sería el metadato.

En un ámbito más familiar para nosotros, como es la biblioteca, se puede dar este caso.

En un libro de Pío Baroja (dato), el nombre: Baroja, Pío es un metadato y la categoría Autor principal es el meta-metadato.

Pero, si consideramos el ámbito del catálogo de esa biblioteca, entonces "Baroja, Pío" es el dato, "Autor principal" sería el metadato y la categoría "Índice de autores" sería meta-metadato.

En cualquier contexto particular, los metadatos deben poseer un nivel de abstracción superior que el dato que describen. El metadato es una categoría lógica superior en la que se incluye el dato.

Normalmente no es posible distinguir entre datos "puros" y metadatos pues:

- Se puede ser dato y metadato al mismo tiempo. El título de un artículo es tanto título (metadato) como parte del texto (dato).
- Datos y metadatos pueden cambiar sus funciones. Un poema como tal sería considerado como dato, pero si existiese una canción que lo usara como letra, el poema podría relacionarse con un archivo de sonido de la canción como metadato. Así, la clasificación depende del punto de vista.
- Es posible crear meta-meta-...-metadatos. Aunque en principio esto parezca inútil, puede ser imprescindible archivar metadatos sobre metadatos, por ejemplo para registrar la procedencia de los metadatos al fusionar dos documentos.

SISTEMAS, ESQUEMAS O MODELOS DE METADATOS

La creación de metadatos requiere el uso de un esquema o modelo. Existen casi tantos como proyectos. Se pueden distinguir diferentes iniciativas de metadatos:

Metadatos de propósito general, dirigidos a cualquier tipo de información, destinados a describir cualquier documento. (MARC, Dublín Core, MODS, METS)

Metadatos de propósito específico, pensados para describir una tipología específica (información geográfica, fotográfica, musical), restringida a un ámbito temático concreto, o a una comunidad informativa (empresa, institución).

TEI. Text Encode Initiative, sobre literatura clásica y sus traducciones

EAD. Encoded Archival Description, para documentos de archivo

IMS. Instructional Management System, para entornos de formación

Metamodelos de metadatos

Permiten codificar distintos modelos de metadatos.

RDF. Metamodelo desarrollado por W3C (Consortio WWW), que utilizando lenguaje XML sirve de estructura común a diferentes modelos de metadatos, haciendo que la información de la web se recupere y comparta de forma más satisfactoria.

Los Metadatos sobre modelos (de metadatos) se llaman metamodelos. RDF

TIPOS DE METADATOS

En cualquier sistema de metadatos

Según su aplicación:

Administrativo: Usados en la gestión y administración de recursos de información

Descriptivo: Usados para representar recursos de información

Preservación: Para salvaguardar y conservar los recursos de información

Técnico: Relativos al funcionamiento de los sistemas o el comportamiento de los metadatos en un sistema dado.

Uso: Relativos al nivel y tipo de uso que se hace de los recursos de información.

En un sistema de metadatos puede haber de una, varias o todas las clases. Por lo general, Administrativo, Descriptivo y Técnico.

Según su almacenamiento

Los metadatos se pueden almacenar *internamente*, en el mismo archivo que el dato, o *externamente*, en un archivo separado.

(Metadatos internos)

1. Información incluida en el código fuente del propio recurso electrónico.

En el entorno web, de momento, hay pocos recursos que contengan metadatos fiables. Los buscadores más desarrollados no los han tenido en cuenta, hasta ahora, para sus búsquedas, sino que utilizan complejos sistemas de reconocimiento de palabras o secuencias, o bien en texto completo. El futuro va en la dirección de crear documentos web con metadatos fiables, normalizados y asimilables en diversos idiomas y modelos (Web semántica) que se convertirían en accesos para los programas de búsqueda.

(Metadatos externos)

2. Información contenida en bases de datos, catálogos, repositorios. Sirven como acceso desde los programas de búsqueda internos, con varios sistemas de enlace a la información referida.

PRINCIPALES SISTEMAS DE METADATOS PARA LA DESCRIPCIÓN BIBLIOGRÁFICA

FORMATO MARC (ISO 2709)

MARC es el acrónimo de Machine Readable Cataloging y es un sistema o modelo de metadatos que estructura la información contenida en un registro de metadatos. Es un formato internacional de almacenamiento y transferencia de registros.

Se estructura en tres partes bien diferenciadas: cabecera, directorio y campos de datos.

Cabecera: contiene 24 elementos codificados de longitud fija que proporcionan al sistema de gestión, información de control sobre el conjunto del propio registro, como su longitud, nivel de descripción, etc. Sirven para identificar el registro y distinguirlo de otro de manera unívoca. En muchos casos, su información son meta-metadatos.

Directorio: es un índice generado por el ordenador acerca de los campos incluidos en el registro y no se visualiza, pues contiene información para el propio sistema.

Campos de datos: 11 elementos de descripción, identificados por una etiqueta de campo que codifica cada área de información bibliográfica. La etiqueta de campo está formada por 3 dígitos entre 000 y 999. A continuación, los indicadores, 2 posiciones con valores numéricos o vacíos, son calificadores del metadato/campo principal. Sirven para especificar características más definidas dentro del esquema MARC. Las combinaciones de los dígitos de la etiqueta y los indicadores incrementan el número de elementos de manera significativa.

Otros elementos que sirven para calificar la información son los subcampos, que sirven para identificar partes de la información de un metadato o para delimitar su área de aplicación.

Tiene las ventajas e inconvenientes de un modelo tradicional

Ventajas:

- Estable y bien mantenido
- Buena interoperabilidad y comprensión en su entorno
- Proporciona descripciones de contenido complejas, con precisión en la recuperación
- Gran comunidad de usuarios

Inconvenientes:

- Dificultad para ser flexible o extensible
- Adaptación a entornos concretos limitada
- Muy complejo
- Poco uso fuera del mundo bibliotecario

MARC está muy implementado como modelo de metadatos en bibliotecas. Actualizado y mantenido por la Network Development and MARC Standards Office de la Library of Congress. Sigue de cerca el orden y la estructura de ISBD y es compatible con las Reglas de catalogación (AACR, y todas las europeas) y por eso, se utiliza en la gran mayoría de los sistemas de gestión bibliográfica, por lo que la interoperabilidad de la información está garantizada. De hecho, existen importantes catálogos colectivos a todos los niveles (WorldCat, REBIUN, CCPB) y se pueden captar y reutilizar registros creados y almacenados en distintos sistemas porque los metadatos en MARC son intercambiables.

Sin embargo no está diseñado especialmente para información menos bibliotecaria, información digital web, principalmente, aunque de hecho se utiliza habitualmente en la descripción de los materiales no librarios de las bibliotecas.

Es un modelo bastante complejo, con gran cantidad de elementos y con un alto nivel de codificación. Un registro MARC puede contener una gran cantidad de información, tanto codificada mediante los códigos MARC para lenguas, países, tipologías documentales, materiales, tamaños, etc., (006, 007, 008), como en lenguaje natural o controlado (Reglas de catalogación, Tesoros).

Los elementos/campos MARC se aplican a distintas funciones:

MARC Bibliographic data, metadatos de descripción

MARC Authority data, metadatos de recuperación y autenticación de la información

MARC Holdings data, metadatos de uso y gestión

DUBLIN CORE (DCMI . Dublín Core Initiative Metadata) ISO 15836-2003

Es el modelo o esquema de metadatos más utilizado en el entorno Web y sus elementos pueden convertirse en la infraestructura para la construcción de la Web semántica. Tiene estas características, propias de un sistema menos tradicional.

Ventajas:

- Muy flexible y extensible
- Fácil adaptación a entornos concretos
- Más utilizado y comprendido en entornos no bibliotecarios
- Admite diversos grados de complejidad

Desventajas:

- En su entorno funcionan otros modelos de metadatos
- Todavía no están resueltos los problemas de interoperabilidad entre los distintos modelos
- No suele utilizarse en el ámbito bibliotecario común

DC parte de un modelo simple de 15 elementos básicos suficientes para describir cualquier documento digital (DLO) y aparecen divididos en 3 grupos que indican la clase o alcance de la información contenida en ellos.

Contenido:

Title/Título

Subject/Materia

Description/Descripción (Resumen)

Source/Fuente

Language/Idioma

Relation/Relación

Coverage/Cobertura

Propiedad intelectual:

Creador/Creador

Publisher/Editor

Contributor/Participante

Rights/Derechos

Características:

Date/Fecha

Type/Tipo

Format/Formato

Identifier/Identificador

Además de los elementos básicos (no obligatorios y todos repetibles) DC puede adaptarse a necesidades de descripción más específicas mediante unos mecanismos:

-Perfiles de aplicación, asociados a distintas disciplinas.

-Términos de metadatos, donde pueden incluirse nuevos elementos que se añaden al esquema básico, matizaciones de los elementos y también términos de vocabularios controlados que se utilizan para crear los metadatos.

INTEROPERABILIDAD

El problema de la interoperabilidad entre los distintos modelos de metadatos está aún sin resolver completamente, aunque hay herramientas que contribuyen a que sea más factible, como el lenguaje de marcado XML (ISO 8879) es una estructura de lenguaje de marcas que facilita el intercambio de datos bibliográficos a través de la web.

MARC ha creado distintos modelos en XML, unos más complejos que otros como:

MARCXML, Metadata Object Description Schema (MODS), Metadata Encoding & Transmission Standard (METS).

El futuro va por establecer las equivalencias entre los distintos esquemas o estándares de descripción para que puedan ser compatibles e interoperar.

El metamodelo RDF (Resource Description Framework), es una infraestructura de descripción de recursos, que permite codificar en elementos comunes diferentes modelos de metadatos. Propone un modelo coherente y un marco sintáctico para representar términos de metadatos e identifica mediante URI (Uniform Resource Identifiers) la ubicación de los esquemas donde se declaran y definen los términos.

La utilización de pasarelas (cross-walks) a través de tablas de equivalencias (mapping), también es clave para la transmisión de la información aunque esté en modelos diferentes. Ya existe un desarrollo MARC-XML, así como un formato de conversión o pasarela MARC-DC-GILS (modelo de metadatos de información geográfica).

DISEÑO, EVALUACIÓN, MANTENIMIENTO

Metadatos de biblioteca (mixta o digital)

Hay tres categorías de metadatos que se usan frecuentemente para describir objetos en una biblioteca

- 1- **Descriptivos**- Información que describe el contenido intelectual del objeto, ayuda a la búsqueda o esquemas similares. Se usa típicamente con propósitos bibliográficos y para búsqueda y recuperación.
- 2- **Estructurales**- Información que une cada objeto a otros para formar unidades lógicas (p.ej. información que relaciona partes individuales de un documento a las otras que lo componen). Se utilizan mucho en la descripción de documentos digitales
- 3- **Administrativos**- Información usada para gestionar el objeto o controlar el acceso a él. (metadatos acerca de la adquisición, vigencia y préstamo, por ejemplo). En las bibliotecas digitales puede incluir información sobre cómo fue escaneado, su formato de almacenamiento, información de copyright y licencias, e información necesaria para la preservación a largo plazo de los objetos digitales.

Para crear y desarrollar un sistema de información digital hay varias opciones:

- Utilizar un esquema o modelo existente
- Crear un modelo propio
- Seleccionar un modelo y adaptarlo a nuestras necesidades

Como regla general es mejor utilizar un modelo ya existente, adecuado a nuestro tipo de información; o en todo caso adaptarlo a nuestras necesidades. De esta manera se ahorra una gran cantidad de tiempo y esfuerzo. En caso de crear un modelo propio, merece la pena crearlo con las herramientas y condiciones que nos aseguren la mayor difusión e interoperabilidad.

Hay que tener en cuenta estas consideraciones:

1. El tipo de información digital que queremos describir (texto, música, audio/video), o sus particularidades (material de archivo, material de enseñanza)
2. El tipo de proyecto para el que vamos a elegir el sistema puede ser:
 - Una colección digital formada por documentos tradicionales digitalizados y/o por documentos digitales originales. En este caso el control se ejerce sobre los documentos y sobre los metadatos.
 - Una colección virtual constituida por un conjunto de metadatos que describen y caracterizan el acceso a documentos accesibles de índole diversa. En este caso, el control sólo se ejerce sobre los metadatos.
3. Si existen metadatos previos sobre los documentos, tanto de un modelo estructurado (MARC, TEI), como metaetiquetas incluidas en la cabecera de un documento originalmente digital. Es importante elegir un modelo que pueda reutilizar fácilmente los

- metadatos preexistentes.
4. Si los metadatos van a ir incluidos en el propio documento, o bien van a formar una base de datos separada con el enlace correspondiente al documento.
 5. El personal destinado, teniendo en cuenta su número, formación, perfil profesional (bibliotecarios, informáticos, especialistas en diversos temas...)
 6. La recuperación de la información que se desea establecer, teniendo en cuenta la necesidad mayor o menor de normalización y el nivel de interoperabilidad con otros sistemas o colecciones semejantes, bien para crear proyectos conjuntos (catálogos colectivos, repositorios comunes), bien para aprovechar metadatos de otros proyectos o colecciones.

El fin de la descripción es hacer más comprensible la información, así pues, al evaluar los sistemas de descripción bibliográfica deberemos tener en cuenta:

Su tipo de codificación. Actualmente cualquier formato o modelo debería codificarse en XML. Los modelos de metadatos son necesarios para la descripción eficiente.

El control del significado de los metadatos debe estar asegurado por el uso de esquemas de contenido, como Tesauros, listados, códigos, etc. (Las famosas listas desplegadas para elegir un valor de contenido).

La fiabilidad de la descripción se basa, tanto en la adecuación del modelo utilizado, como de la responsabilidad humana implicada. Las instituciones que respaldan una y otra son quienes autentifican la descripción.

Finalmente, las cuestiones relativas a la recuperación y movilidad o transporte, se confirman mediante su adecuación y compatibilidad en protocolos de transmisión y presentación.

Las tareas implicadas en el mantenimiento de los sistemas de metadatos son complejas y abarcan varios aspectos:

Mantenimiento de hardware en general (servidores, ordenadores, conexiones y cableado)

Mantenimiento de software (sistemas o programas de gestión documental, protocolos)

Mantenimiento de los modelos de metadatos utilizados, mediante la puesta al día del esquema de estructura (los elementos del modelo). Los modelos de metadatos más utilizados tienen actualizaciones y desarrollos periódicamente.

Mantenimiento de los esquemas de contenido de los metadatos, mediante la actualización y revisión de sus contenidos (si un país aparece/desaparece o cambia de nombre, la evolución orgánica de una institución, los términos de los tesauros o de las listas de materias utilizadas, los códigos de ubicación de las colecciones, etc.)

FUENTES DE INTERÉS

Curso de Madroño sobre metadatos:

<http://rayuela.uc3m.es/~gbueno/madrono/presentacion.html>

Curso de SEDIC sobre metadatos: <http://www.sedic.es/autoformacion/metadatos/>

Web de Dublín Core: <http://dublincore.org>

Web de MARC: <http://locgov/marc>