

Organización de la información con RDA: su presencia en los catálogos de bibliotecas de América Latina



Filiberto Felipe Martínez Arellano
Silvia Mónica Salgado Ruelas
Patricia de la Rosa Valgañón
Coordinadores



Z666.5

O74

Organización de la información con RDA : su presencia en los catálogos de bibliotecas de América Latina / Coordinadores Filiberto Felipe Martínez Arellano, Silvia Mónica Salgado Ruelas, Patricia de la Rosa Valgañón. -- México : UNAM. Instituto de Investigaciones Bibliotecológicas y de la Información : UNAM. Instituto de Investigaciones Bibliográficas : UNAM. Dirección General de Bibliotecas, 2020.

421 p. -- (Organización de la Información)

ISBN: 978-607-30-3621-4

1. Organización de la información -- Bibliotecas -- América Latina 2. RDA (Recursos, descripción y acceso) -- 3. Metadatos -- Bibliotecas -- América Latina. 4. Catálogos -- Bibliotecas -- América Latina I. Martínez Arellano, Filiberto Felipe, coordinador. II. Salgado Ruelas, Silvia Mónica, coordinadora. III. Rosa Valgañón, Patricia de la, coordinadora. IV. ser.

Diseño de cubierta:

José Miguel García Fernández

Primera edición, 2020

D.R. © UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Ciudad Universitaria, 04510, Ciudad de México

Impreso y hecho en México

ISBN: En trámite

Publicación dictaminada 2020

La web semántica y organización de información

JUAN FRANCISCO MORENO MEJIA

BIBLIOTECA NACIONAL PEDRO HENRÍQUEZ UREÑA, REPÚBLICA DOMINICANA

INTRODUCCIÓN

En los nuevos tiempos, en que predomina el uso de las tecnologías para el acceso y difusión del aprendizaje y el conocimiento, se hace perentorio disponer de las herramientas adecuadas que permitan una gestión efectiva y eficiente en la organización de información. En la elaboración de este documento, abordamos los metadatos y sus funciones en la organización de la información para que ésta pueda ser localizada, identificada, accesada y utilizada de forma efectiva; estudiamos los esquemas de metadatos, sus roles, su tipología y su ciclo de vida.

Definimos el concepto de web semántica, su arquitectura y su sintaxis, la importancia de ésta en el acceso y difusión de la información, y finalmente realizamos una comparación de web sintética versus web semántica, ventajas y desventajas.

METADATOS

El término “metadatos” no tiene una definición única. La definición más difundida es que son “datos sobre datos”. También hay otras como “informaciones sobre datos” (Sheldon 2001) «datos sobre informaciones» (Ghavam y Steinmetz 2001) e «informaciones sobre informaciones» (Swick 2002). Son ejemplos de metadatos el resumen de un documento, los registros catalográficos en cualquier formato (MARC, Dublin Core), una dirección IP, el índice de un documento, la extensión de un documento multimedia, los datos que nos ofrece el ordenador sobre cualquier documento (fecha y usuario de creación/modificación, peso del fichero), etcétera.

De acuerdo con la UNE-ISO (23081-1, 2008), los metadatos son “información estructurada o semi estructurada que posibilita la creación, registro, clasificación, acceso, conservación y disposición de los documentos a lo largo del tiempo”.

Los metadatos incluyen una amplia información que se puede utilizar para identificar, autenticar y contextualizar los documentos, las personas, los procesos de negocio la regulación y sus relaciones.

Ciclo de vida de los metadatos

Creación, manipulación y destrucción de los metadatos

Se pueden crear metadatos manualmente, semiautomáticamente o automáticamente. El proceso manual puede ser muy laborioso, depende del formato usado y del volumen deseado, hasta un grado en el que los seres humanos no puedan superarlo. Por eso, el desarrollo de instrumentos semiautomático o automático es más recomendable.

En la producción automática, el software adquiere las informaciones que necesita sin ayuda externa; en la producción semiautomática un servidor humano sostiene algoritmos autónomos con la aclaración de inseguridades o la proposición de informaciones que el software no puede extraer sin ayuda.

Funciones de los metadatos

Según Taylor (2004), las funciones primarias de los metadatos son facilitar la identificación, ubicación, recuperación, manipulación y uso de recursos de información accesibles en línea.

Tipología

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

1. Descriptivos: información bibliográfica.
2. Estructurales: referidos a formatos y estructuras.
3. Administrativos: incluye permisos, copyright y toda información relacionada al acceso.

Según Weibel (1997), existen tres modelos para la creación y uso:

1. Modelo embebido, donde los metadatos son formulados por el creador del documento electrónico y la descripción viene incluida en el mismo. Esto

equivale a la catalogación en la fuente o catalogación en publicación (CIP).

2. Generados por terceros, donde una agencia crea, reúne y administra los registros, pero no están incluidos en el registro en sí. Se generan productos similares a las bases de datos referenciales.
3. Visualización filtrada, en este modulo una agencia administra registros de metadatos, en una descripción común compartida por todos los usuarios.

Tiene la ventaja de proveer una interface única de búsqueda. De acuerdo con la riqueza de la descripción y nivel de detalle, se dividen en tres niveles (Dempsey & Heery, 1997):

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

Nivel 2: Estructurados, son relativamente nuevos y muchos han sido diseñados *ad hoc* para su uso en internet. Inicialmente no tenían normas, pero actualmente algunos de ellos están surgiendo como normas, soportan búsquedas por campos y son relativamente fáciles de aplicar, están pensados para que los incluya el autor del documento; por ejemplo, Dublín Core.

Nivel 3: Excelente 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; por ejemplo, el formato MARC.

Modelos de metadatos para la gestión de documentos

Las instituciones u organizaciones deberían definir los requisitos para diseñar e implementar los metadatos para la gestión de documentos y como estos interactúan con los sistemas de información. Estos requisitos deberían establecerse a partir de la identificación del entorno legal y reglamentario, los riesgos asociados a los procesos y las necesidades.

El diseño y la implementación de metadatos, conlleva a que las instituciones definan los roles y responsabilidades de las personas y establezcan un esquema de metadatos.

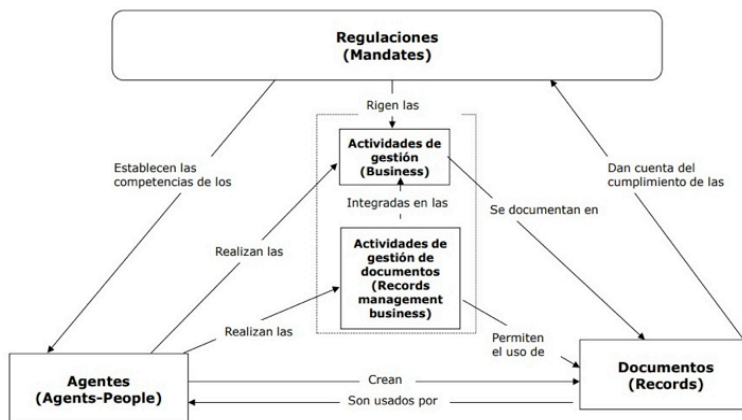
Roles y responsabilidades

Los roles y las responsabilidades frente a la creación, el uso, la implementación y el mantenimiento de los metadatos para la gestión de documentos deberían definirse, asignarse y comunicarse a todo nivel de la institución. Dentro de la definición de los roles y responsabilidades se deberían tener en cuenta lo siguiente:

1. Responsables de la gestión de documentos. Deben ser los encargados de la confiabilidad, autenticidad, disponibilidad e integridad de los metadatos asociados a los documentos.
2. Responsables de las tecnologías de información. Son responsables de la confiabilidad e integridad de los sistemas utilizados para crear y mantener los metadatos.
3. Los directivos. Son responsables de que los controles internos funcionen adecuadamente.

4. Los técnicos (catalogadores). Son los responsables de crear los metadatos y de que sean precisos y estén completos.

Figura 1. Metadatos para la gestión de documentos



Fuente: UNE ISO – 23081-1 y 2.

ESQUEMAS DE METADATOS

El objetivo de un esquema de metadatos para la gestión de documentos es mostrar de manera lógica las relaciones entre los diferentes componentes del conjunto de metadatos a través de reglas para el uso y la gestión específicamente relacionadas con la semántica, la sintaxis y la obligatoriedad de los valores (Asociación Española de Normalización)

Análisis de los esquemas de metadatos

Aspectos seleccionados:

Fecha de creación, institución creadora, objetivos, corriente teórica a la que responden, simplicidad/complejidad, interoperabilidad y nivel de catalogación.

Dublin Core

Es un esquema de metadatos que fue creado a partir de 1995 por Online Computer Library Center (OCLC), el cual fue concebido con el objetivo de la creación de recursos electrónicos para ser consultados en Internet. Este modelo responde a la corriente de informáticos y bibliotecarios en cuanto a su simplicidad o complejidad. Podemos decir que es un sistema de mediana complejidad que puede utilizarse con calificadores o sin ellos, lo que lo caracteriza como un sistema sencillo y amigable al usuario, utiliza un nivel de catalogación nivel 2 y su estructura está compuesta por quince elementos divididos en tres grandes grupos.

MODS

Es un esquema creado por la Library of Congress (LC) en 2002. El objetivo de su creación es formar parte de un subconjunto de MARC21. Responde a la corriente bibliotecaria. En cuanto a la simplicidad o complejidad es un modelo más compatible con las bibliotecas; sin embargo, tiene cierto nivel de complejidad, por lo que se considera con una complejidad mediana y está orientado al usuario final.

Este modelo puede ser exportado a diferentes formatos; no obstante, el esquema no apunta a la redondeabilidad con MARC21. En otras palabras, un registro MARC21 original convertido a MODS no puede convertirse de nuevo a MARC21 en su totalidad sin cierta pérdida de especificidad en el marcado o pérdida de datos. El nivel de catalogación es bastante sencillo, es más orientado al usuario que MARC XML. Permite la conversión de los campos centrales de MARC21, aunque podrían perderse algunos datos.

METS

METS es un estándar para codificar metadatos descriptivos, administrativos y estructurales relacionados con objetos dentro de una biblioteca digital. Se basa en los esquemas del lenguaje XML, fue creado por Digital Library Federation y es mantenido por la Biblioteca del Congreso. Su objetivo es la gestión de metadatos en bibliotecas digitales, pero su función principal es el intercambio de archivos audiovisuales que contengan imágenes, videos o sonidos.

Podemos decir que pertenece a la corriente teórica de bibliotecarios. En cuanto a su simplicidad o complejidad, es un sistema complejo, a pesar de que fue concebido para promover la interoperabilidad de los contenidos digitales entre los sistemas de bibliotecas digitales y preservación de los materiales de las bibliotecas digitales. Utiliza un nivel 3 de catalogación.

MARC21/MARCXML

Es un modelo de metadatos creado por Library of Congress (LC) en el año 1960. En 1992 se incluyó el campo 856 y la

versión inicial de MARCXML en el 2002. Su mantenimiento está a cargo de Network Development and MARC Standards Office, Library of Congress (Desarrollo de Redes y Normas MARC, Biblioteca del Congreso). Su objetivo fue inicialmente para la catalogación de documentos impresos, especialmente obras monográficas.

La corriente teórica proviene de la comunidad bibliotecaria, es un modelo bastante complejo para la creación de un registro bibliográfico completo, se requiere personal especializado y capacitado que tenga pleno conocimiento de las reglas de catalogación. Es un sistema flexible para la interoperabilidad, puede exportar a distintos formatos bibliográficos y de metadatos, tiene catalogación nivel 3 (Demsey 1997). Aunque posee un nivel mínimo para Estados Unidos, permite mayor nivel de detalle empleando campos opcionales o subcampos.

Reflexión del análisis

Los cuatros modelos analizados tienen común denominador, que es su interoperabilidad, lo que permite el intercambio de registros de metadatos; sin embargo, no todos poseen la misma flexibilidad. Ejemplo de esto es que el esquema de MARC21 permite la conversión al modelo MODS, pero éste no puede ser regresado de nuevo a MARC21 en su totalidad, debido a la pérdida de especificidad en el marcado o pérdida de datos.

Los esquemas MARC21/MARCXML y MODS tienen la particularidad de que fueron creados por la misma institución y se puede considerar que el primero es un derivado del segundo, el cual fue creado con la finalidad de extraer datos de manera selectiva de los registros MARC21 existentes.

Ambos modelos fueron diseñados para la descripción de documentos en distintos formatos, los cuales son utilizados especialmente por la comunidad bibliotecaria.

El esquema Dublin Core es el modelo más adecuado para el manejo de objetos digitales, y le permite a los autores describir sus propios recursos web, para que puedan ser recuperados por los motores de búsqueda de Internet.

Los METS fueron diseñados para la gestión de metadatos de bibliotecas digitales, pero su fuerte radica en el intercambio de archivos audiovisuales que contengan imágenes, videos o sonidos.

WEB SEMÁNTICA

El término “semántico” se refiere a los aspectos del significado, sentido o interpretación de signos lingüísticos como símbolos, palabras, expresiones o representaciones formales. En principio las expresiones del lenguaje formal o de una lengua natural admiten algún tipo de correspondencia con situaciones o conjuntos de cosas que se encuentran en el mundo físico o abstracto que puede ser descrito por dicho medio de expresión.

La web semántica (del inglés *semantic web*) es un conjunto de actividades desarrolladas en el seno del World Wide Web Consortium con tendencia a la creación de tecnologías para publicar datos legibles por aplicaciones informáticas (máquinas en la terminología de la Web semántica). Se basa en la idea de añadir metadatos semánticos y ontológicos a la W3C. Esas informaciones adicionales que describen el contenido, el significado y la relación de los datos, se debe proporcionar de manera formal para que así sea posible evaluarlas automáticamente por máquinas de procesamiento.

Podemos decir que la Web semántica es un conjunto de iniciativas destinadas a convertir la W3C en una gran base de datos capaz de soportar un procesamiento sistemático y consistente de la información.

El precursor de la idea, Tim Berners-Lee, intentó desde el principio incluir información semántica en su creación, la World Wide Web, pero por diferentes causas no fue posible (Carvin 2005). Por ese motivo introdujo el concepto de semántica con la intención de recuperar dicha omisión.

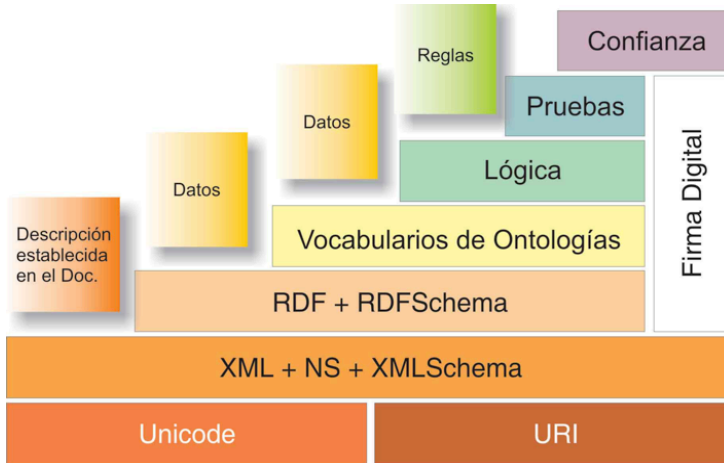
Cuando se realiza una búsqueda en la web, los navegadores realizan esta actividad en bases de datos especializadas que constantemente son actualizadas por web, una especie de robot que ejecuta consultas constantes entre enlaces de hipertexto para añadir la información que encuentran a sus bases de datos. Con esta operación proactiva agilizan la entrega de resultados y mantienen actualizadas las diferentes entradas, ya sea eliminándolas cuando el sitio se ha retirado o modificándolas cuando sea necesario. Esto se debe a que la WWW es un entorno dinámico, que cambia constantemente según las necesidades de sus propietarios.

ARQUITECTURA

DE LA WEB SEMÁNTICA Y LA VINCULACIÓN DE DATOS

La Web Semántica va más allá de poner datos en la web. Implica establecer relaciones (relaciones de *significado*) que permiten conectar recursos y enlazarlos. La web se entiende, por tanto, como una gran nube de datos enlazados. Pero para lograr esto es necesario desarrollar toda una arquitectura, compuesta por una serie de componentes y capas sobre la base de la Web.

Figura 2. Modelo de capas propuesto por Berners-Lee para la web semántica



Descripción de las capas que forman la arquitectura de la Web semántica

Universal Resource Identifier URI. Es una cadena formateada que sirve como medio para identificar la ubicación de recursos físicos o abstractos.

Unicode. Provee un número único para los caracteres, independientemente de la plataforma, programa o lenguaje subyacente, es requerido para estándares tales como XML, Java, LDAP, WML, etcétera.

XML + NS + XMLSchema, es el lenguaje XML que permite separar el contenido de la presentación de formatos de datos, y proporciona la sintaxis superficial para los documentos estructurados. XMLSchema es un lenguaje que permite definir la estructura de los documentos XML.

RDF + RDFSchema. RDF provee un modelo para describir recursos en la web, que son identificables únicamente a través de una URI y que describen la información de manera que pueda ser entendida por las aplicaciones de los computadores. RDFSchema es el vocabulario que permite describir las propiedades y las clases de los recursos RDF; además, aporta la semántica para establecer jerarquías de generalización entre las propiedades y clases.

RDF + RDFSchema provee un modelo para describir recursos en la web que son identificables únicamente a través de una URI y que describen la información de manera que pueda ser entendidas por las aplicaciones de los computadores.

Vocabularios de ontologías. Es un lenguaje para construir ontologías a través de la descripción detallada de propiedades y clases, tales como relaciones entre clases, cardinalidad, igualdad, tipologías de propiedades más complejas y caracterización de propiedades o clases enumeradas.

Logic, Proof y Trust. Son capas que han sido explotadas significativamente.

Firma Digital. Es utilizada en las capas 6, 7 y 8 como componentes de seguridad.

Diferencias entre web sintáctica y web semántica

Web sintáctica	Web semántica
<ul style="list-style-type: none">• Se sirve de las palabras y los términos para ser localizadas por los buscadores.• Almacena, prácticamente toda la información del mundo.	<ul style="list-style-type: none">• Web extendida, está centrada en el significado.• Dota de significado al contenido.

Web sintáctica	Web semántica
<ul style="list-style-type: none"> • Presenta un acceso casi instantáneo a la información desde cualquier lugar del planeta con una conexión a Internet. • La información disponible es textual, documentos formados por imágenes, videos, presentaciones, etcétera. • Conjunto de recursos enlazados entre si formando un grafo dirigido. • Numerosos buscadores enlazan con la información para movernos por ella. 	<ul style="list-style-type: none"> • Procesamiento de grandes volúmenes de contenido. • Optimiza el tratamiento de dichos contenidos. • Se obtienen respuestas rápidas y sencillas. • La información se obtiene de forma definida y concreta. • Los problemas tienen una infraestructura común. • La web semántica permite compartir, transferir y procesar archivos. • Usa metalenguajes como XML, XML Schema, RDF Schema y OWL.
Ventajas de la web sintética	Ventajas de la web semántica
<ul style="list-style-type: none"> • Cualquier persona puede añadir más información a la Web • Busca y hace uso de variada información. • Presenta acceso a la información desde cualquier lugar del planeta. • Pone a disposición todo tipo de información (textual, documentos formados por imágenes, videos, presentaciones, etcétera). • Recopila prácticamente toda la información del mundo. 	<ul style="list-style-type: none"> • Facilita la búsqueda de información. • Permiten la gestión del conocimiento, el razonamiento automático y la búsqueda de información de manera precisa y eficaz. • Con esta información podemos realizar consultas complejas en buscadores. • Permite el trabajo cooperativo entre las personas y los computadores. • Permite una mejor organización de la información. • Se apoya en lenguajes universales.

Desventajas de la web sintáctica	Desventajas de la web semántica
<ul style="list-style-type: none">• Problemas de interoperabilidad de aplicaciones, que se deben a la falta de entendimiento sintáctico o a la incapacidad de un sistema de información para leer datos de otros sistemas.• Ralentiza la implantación de nuevos servicios útiles.• Repercute en el costo de los servicios.• A pesar de los buscadores y de su potencia, las respuestas no son adecuadas a las preguntas que realizan los usuarios.• Los buscadores no enlazan con la totalidad de páginas existentes.• Devuelven millones de resultados generando escasa precisión de los resultados.• El documento buscado debe estar descrito con las mismas palabras que se introducen al buscar.• Deben actualizar permanentemente los índices a través de robots.• Necesita de la Web semántica para tratar de resolver todos estos problemas.	<ul style="list-style-type: none">• Problemas de interoperabilidad de aplicaciones.• Costoso y laborioso proceso de adaptar y reestructurar los documentos de Internet.• Problemas del idioma.• Complejidad de la codificación semántica.• Necesita unificar los estándares semánticos.

CONCLUSIÓN

Después de realizar el estudio de los componentes que integran la web semántica, podemos decir que la web semántica se ha convertido en el paradigma de las plataformas de acceso y difusión de la información, al añadirle metadatos semánticos y ontológicos a la (W3C). La web semántica es un conjunto de iniciativas destinadas a convertir la W3C en una gran base de datos capaz de soportar un procesamiento sistemático y consistente de la información, lo que permite que los usuarios puedan obtener información de forma ágil y precisa.

Se investigaron los principales elementos que interactúan en la organización de información. Entre éstos se encuentran los metadatos, sus modelos y esquemas. Su importancia se determina en la creación, el registro, la clasificación, el acceso, la conservación y la disposición de los documentos. Existen diversas definiciones de metadatos, aquí empleamos la más corta de todas: “datos sobre datos”.

REFERENCIAS

- Archivo General de la Nación Colombia. Guía de metadatos: guía para la formulación de un esquema de metadatos para la gestión de documentos. https://www.archivogeneral.gov.co/sites/default/files/Estructura_Web/5_Consulte/Recursos/Publicacionees/GuiaDeMetadatos.pdf.
- Asociación de Archiveros y Gestores de Documentos. <http://www.arxivvers.com/index.php/documents/documentacio-1/normativa-tecnica-1/507-tec-isotc46sc11-n800r1-orienta-elab-esq-metad-es-1/file>.

- Asociación Española de Normalización. Información y documentación. Procesos de gestión de documentos. Metadatos para la gestión de documentos. Parte 1: Principios. UNE-ISO 23081-1. Madrid: AENOR, 6.
- Blázquez Ochando, Manuel (2012). Sistema de Recuperación e Internet. Disponible el 22 de agosto de 2017 en <http://ccdodoc-sistemasrecuperacioninternet.blogspot.com/2012/11/metadatos-para-la-descripcion-de.html>.
- Carvin, Andy (2005). "Tim Berners-Lee: Weaving a Semantic Web". *Digital Divide Network*.
- Dempsey, L y Heery, R (1997). *Specification for resource description methods*, Parte 1. A review of metadata: a survey of current resource description formats.
- Dfutura Soluciones Multimedia (2015). "La web 3.0 o web inteligente, el futuro ya está aquí". Disponible el 1 de agosto de 2017 en <https://www.dfutura.com/blog/la-web-3-0-o-web-inteligente-el-futuro-ya-esta-aqui>.
- Guía para la formulación de un esquema de metadatos para la gestión de documentos del archivo general de la nación de Colombia. Disponible en el portal de Editur, <http://www.editeur.org/83/Overview/>.
- Ghavam, Steinacker, y Steinmetz (2001). Metadata Standards for Web-Based Resources. IEEE
- MultiMedia. <http://www.dsc.ufcg.edu.br/~garcia/cursos/TEICOPIN/metadataWE.pdf>.
- La Web Sintáctica vs. La Web Semántica. Disponible el 29 de agosto de 2017 en <https://sites.google.com/site/websemanticapatty/la-web-sintactica-vs-la-web-semantica>.
- Lamarca Lapuente, María Jesús. Hipertexto: El nuevo concepto de documento en la cultura de la imagen. La web Semántica. http://www.hipertexto.info/documentos/web_semantica.htm.

- Lazinger, Susan S. (2001). *Digital Preservation and Metadata*. Englewood, Colorado: Libraries Unlimited.
- Library of Congress. METS, Metadata Encoding Transmission Standard Disponible el 28 de agosto de 2017. <http://www.loc.gov/standards/mets/>.
- _____. MODS, Metadata Object Description Schema. Disponible el 28 de agosto de 2017. <http://www.loc.gov/standards/mods/>.
- McDonough, Jerome P. METS. Standardized Encoding for Digital Library Objects. <https://www.ideals.illinois.edu/bitstream/handle/2142/177/METS.pdf?sequence=2>.
- Metadatos. <https://www.infor.uva.es/~sblanco/Tesis/Metadatos.pdf>.
- Robles Tacury, Andrea Abigail (2014). Web 2.0 y Web 3.0. <https://es.slideshare.net/bglandrea891/ventajas-desventajas-web-30-y-20-34237924>.
- Sheldon, Tom y Big Sur Multimedia (2001). Metadatos. Disponible 29 mayo de 2006 en <http://www.linktionary.com/m/metadata.html>.
- Taylor, A. (2004). *The Organization of information*. Englewood, CO: Libraries Unlimited.
- UNE-ISO/TS 23081-2, p. 12 Principales entidades y sus relaciones.
- W3C, Swick, Ralph (2002). Metadata Activity Statement. Disponible el 29 mayo de 2006 en <https://www.w3.org/Metadata/Activity.html>.
- Weibel, S., Kunze, J., y Lagoze, C. (1997) Dublin Core Metadata for simple resource description.
- Wikipedia. La enciclopedia libre. Metadatos. <https://es.wikipedia.org/wiki/Metadato>.

Organización de la RDA: su presencia en los catálogos de bibliotecas de América Latina.

Instituto de Investigaciones Bibliotecológicas y de la Información, Instituto de Investigaciones Bibliográficas, Biblioteca Nacional y Dirección General de Bibliotecas. La edición consta de 100 ejemplares. Coordinación editorial, Israel Chávez Reséndiz; revisión especializada, Valeria Guzmán González; revisión de pruebas, Carlos Ceballos Sosa; formación editorial, Óscar López Marín. Fue impreso en papel cultural de 90 gr. en los talleres de Grupo Fogra. Año de Juárez 223. Col. Granjas San Antonio. Alcaldía Iztapalapa. Ciudad de México. Se terminó de imprimir en 2020.