

## LOS METADATOS EN INTERNET

Por: Licda. Deyanira Sequeira Ortiz

### 1. ¿Qué son metadatos?

En Internet los metadatos son datos acerca de recursos informativos accesibles en la red. Los bibliotecarios han creado metadatos durante siglos, pero el nombre es nuevo. Los registros bibliográficos construidos utilizando clasificaciones y reglas de catalogación, son esos metadatos. Lo que sucede es que la red *Internet* duplica sus documentos cada medio año, y quienes ponen los insumos y utilizan los diferentes objetos ahí disponibles, han dejado de ser bibliotecarios. Ahora cualquier persona que tenga un computador conectado a la red puede poner su información, lo cual ha producido un caos en la "web". Los usuarios de la red han decidido aprender de los bibliotecarios las formas de organizar la información, construyendo ciertas normas mínimas de convivencia: los metadatos.

Las primeras reuniones de usuarios de la red fueron algo heterogéneas pues participaron especialistas de los más diversos campos del conocimiento: bibliotecarios y archivistas, informáticos, desarrolladores de programas, publicadores, ingenieros en telecomunicaciones, etc. Sus puntos de vista por tanto, también eran disímiles: algunos tenían el punto de vista de los autores y publicadores, otros se planteaban los problemas desde la perspectiva del bibliotecario o archivista, otros se interesaron

por los recursos electrónicos en general y otros por los especializados: en humanidades o geodatos.

Los bibliotecarios se acercaron expectantes a esta nueva realidad, y algunos pensaron que los recursos de Internet debían ser registrados como se hacía con los documentos tradicionales. Sólo que esos estaban constituidos por átomos, ocupaban un espacio, correspondían a un tiempo exacto, a una edición determinada, a un sólo publicador o autor. En tanto que los datos y objetos que maneja Internet son más volátiles: constituidos por octetos, cuya representación simultánea no se refiere más a un espacio o tiempo determinados. En fin, una tecnología nueva que los retaba, probablemente del mismo modo que los primeros incunables retaron a los hombres del Renacimiento, y les permitieron ver más allá de su geografía y de su época.

Hubo un acuerdo acerca de la necesidad de contar con un instrumento muy simple que pudiera ser empleado por todos, y que fuera la base para descripciones más sofisticadas en comunidades especializadas.

## **2. La norma de Dublín.**

La primera reunión, en donde participaron las grandes bibliotecas, se celebró en Dublín, Ohio, y de ahí el nombre de la norma (Dublin Core). Se propuso una hoja de insumo sencilla que permitiera que los publicadores recogieran datos y los incluyeran en sus páginas de *Internet*, y que los desarrolladores de editores para la red pudieran incluir campos para esos datos.

Esta hoja básica permitirá que los sitios de *Internet* dedicados a indizar los objetos, trabajen en forma más eficiente, y que las comunidades especializadas amplíen los datos para adecuarlos a sus necesidades. El grupo se concentró en indicar las características intrínsecas al objeto, y dejó los aspectos exteriores (costo, limitaciones de acceso, formato) para otro momento.

La hoja debe poder ampliarse, del mismo modo que quienes empleamos las hojas de insumo para bases en *microisis*, nos hemos acostumbrado a usar una hoja corta con los elementos básicos, y una hoja larga con muchos más elementos para usuarios especiales o distintos tipos de documentos (audiovisuales, mapas, manuscritos, tesis, proyectos, etc.).

### 3. Elementos incluidos en la norma de Warwick.

Otra reunión para implementar los acuerdos de Dublín, se realizó en Warwick, con la participación de los británicos y de la OCLC. Los aspectos analizados fueron:

- I. La sintaxis: se acordó utilizar una hoja de insumo llamada DTD (documento de definición de tipos, utilizando el lenguaje SGML (lenguaje de marcado generalizado) y las etiquetas usadas en HTML (lenguaje de marcado para hipertextos).
- II. Una hoja de insumo Warwick: que incluyera unos primeros campos esenciales y permitiera ampliaciones a comunidades especializadas, así como inclusión de elementos administrativos, términos, condiciones, etc.

III. Una guía de usuarios: similar a los manuales que empleamos para acompañar las hojas de insumo de BIREME o CEPAL. La norma de Warwick permite que paquetes diferentes, interactúen en la misma hoja. Un paquete es un objeto con tipos específicos de metadatos. Podríamos pensar en el ejemplo de la hoja de insumo de BIREME: los campos del 1 al 4 y el 7, probablemente podrían incluir clasificaciones de Dewey, de LC o NLM, y uno de esos campos podría ser para el URL (la dirección del *dominio* o computador en que se ubica el documento en Internet) y quizás para la descripción en HTML de palabras claves que identifiquen el documento, los derechos de autor, etc. Esa podría ser la forma de conectar los metadatos pertenecientes a paquetes diferentes.

IV. Los profesionales bibliotecarios podrán, a partir de esta norma, incluir aspectos más específicos como calificadores o descriptores que den una descripción más rica de los recursos complejos.

#### 4. Metadatos para imágenes.

Estas reuniones se centraron en el análisis de objetos parecidos a documentos (DLO, document-like objects) o sea, que incluyen texto en su mayor parte. La OCLC convocó a un taller para considerar los recursos visuales: fotografías, slides, y archivos de imágenes, dejando por fuera las imágenes que tuvieran movimiento, las llamadas "animaciones".

El equipo acordó que las normas de las primeras reuniones podían usarse para describir imágenes, pero agregando un elemento para indicar el derecho de autor y el dominio (la computadora donde está el objeto fuente). A partir de ahí, varias bibliotecas y centros informativos han empezado a describir sus objetos: Biblioteca Electrónica de Monticello, la Biblioteca Nacional de Australia, el Servicio Nacional de Documentos e Información de Nueva Zelanda, el Proyecto Nórdico de Metadatos, la Biblioteca Digital de la Biblioteca del Congreso, la OCLC, etc.

#### **5. Implicaciones para Latinoamérica.**

Como resultado de estas reuniones y acuerdos, la LC ha iniciado la traducción de los campos USMARC correspondientes a la norma de Dublín. Pensamos que los bibliotecarios latinoamericanos deberíamos de hacer lo mismo con nuestros formatos para trabajar en *microisis*, por lo que hemos incluido a continuación un breve ejemplo de lo que podría implicar en el caso del formato usado por BIREME. La idea, más que dotar de un instrumento técnico, es hacer pensar a las colegas que recién inician la construcción de sus bases de datos de texto completo, o sus "homepages", que la revolución informativa impulsada en *Internet* tendrá repercusiones en nuestro trabajo que quizás antes no habíamos imaginado, y que estas implicaciones van al fondo mismo de nuestra profesión, porque cambian de manera radical el concepto de información que hemos manejado hasta ahora. En marzo de 1997 se realizará otra reunión en Canberra, Australia en donde se afinará el instrumento de normas. Algunos buscadores

de información como **Altavista e InfoSeek** han empezado a utilizar la norma: al inicio de sus páginas incluyen el METANOMBRE="Tema DC", en donde incluyen las palabras significativas en donde se ha utilizado dicha norma.

Los organismos latinoamericanos que inician su labor pionera de ingreso de información latinoamericana en la red, han de iniciar también este proceso de normalización de etiquetado, para que las búsquedas en Internet se faciliten.

#### **Anexo 1. Hoja de insumo.**

## 6. Referencias.

Workshop Reports. OCLC/NCSA Metadata Workshop.

I. OCLC/NCSA Metadata Workshop Report. Stuart Weibel, Jean Godby, Eric Miller, and Ron Daniel (June, 1995).  
([http://purl.oclc.org/metadata/dublin\\_core\\_report](http://purl.oclc.org/metadata/dublin_core_report)) OCLC/UKOLN Metadata Workshop.

II. The Warwick Metadata Workshop: a framework for the deployment of resource description. Lorcan Dempsey and Stuart Weibel, DLibMagazine (July/August, 1996) (<http://www.dlib.org/dlib/july96/07weibel.html>)

The Warwick Framework: a container Architecture for aggregating sets of metadata. Carl Lagoze, Clifford Lynch, Ron Daniel (June 28, 1996).  
(<http://cs-tr.cs.cornell.edu:80/Dienst/UI/2.0/Describe/ncstr1.cornell%2fTR96-1593>)

CNI/OCLC Workshop on Metadata for Networked Images. Workshop Home Page and Workshop's Executive Summary.  
(<http://purl.oclc.org/metadata/image>)

Australian Digital Library Initiatives. Renato Iannella (editor) (December, 1996). (<http://www.dlib.org/december96/12iannella.html>)

Review of Metadata Formats. Rachel Heery (October 1996).  
(<http://www.ukoln.ac.uk/metadata/review>)

The State of the Dublin Core. Stuart Weibel (January 1997).  
En: International Journal of Digital Libraries, Special Issue on Metadata and Digital Libraries.

Dublin Core Home Page.  
([http://purl.oclc.org/metadata/dublin\\_core](http://purl.oclc.org/metadata/dublin_core)) Metadata Resources (Home Page).  
(<http://www.nlc-bnc.ca/ifla/documents/libraries/cataloging/metadata>)

# ANEXO I

## HOJA DE INSUMO COMPARANDO LAS NORMAS DE BIREME, USMARC Y DUBLÍN

Norma de BIREME	Norma de USMARC	Norma de Dublín
1. Código del Centro 35 ISSN 68. Símbolo 69. ISEN	Identificador. ISBN (020\$a=). Si se usa el URL (856\$b=). Si se emplea el ISSN, entonces (022\$a)	10. <b>Identificador:</b> un grupo de caracteres que dan la identificación única al recurso. Puede ser el URL, o el ISBN
2. Clasificación. 3. Ubicación física 4. Base de datos 5. Tipo de literatura 6. Nivel de tratamiento	Nota sobre restricciones de acceso (506\$a)	15. <b>Derechos de autor:</b> un enlace a alguna nota que indique los derechos de autor.
7. Fuente 30. Serie 53. Conferencia	Entrada de título de la fuente (786\$t)	11. <b>Fuente:</b> el trabajo original de donde se tomó el objeto presente.
16. Autor personal	Nombre personal (100\$a)	2. <b>Creador:</b> la persona u organización primeramente responsable del contenido intelectual del recurso (autores, fotógrafos, ilustradores)
17. autor corporativo	Nombre corporativo (110\$a)	
18. Título	Título propiamente dicho (245\$a)	1. <b>Título:</b> nombre que le da el publicador o creador del documento.
40. Idioma del texto (se usa la norma ISO de dos letras)	Código de lenguaje (041\$a) usando la norma de tres letras.	12. <b>Lenguaje:</b> del contenido intelectual del recurso, que debe coincidir con la norma Z39.53 de tres caracteres, usada para lenguajes escritos.
42. Información descriptiva	<b>Detalle del sistema (538\$a)</b>	9. <b>Formato:</b> la representación del recurso (texto, html, Ascii, archivo postscript, aplicación ejecutable, imagen gif o jpg, etc.). La idea es que quienes usan las máquinas sepan si el recurso se ajusta a su equipo. Hay un listado que incluye tanto los formatos tradicionales (libros, series, etc), como estos nuevos.
43. Tipo de recurso. Este campo ya se usa en las bases SIRPEG del IIDH (Costa Rica)	Tipo de archivo de computador o nota de archivo (516\$f)	8. <b>Tipo:</b> de recurso. La categoría del recurso: si es un homepage, novela, poema, reporte técnico, ensayo, diccionario, etc. Hay una lista de tipos en la siguiente dirección: <a href="http://www.roads.lut.ac.uk/MetadataDC-ObjectTypes.html">http://www.roads.lut.ac.uk/MetadataDC-ObjectTypes.html</a>

61. <b>Notas, y entradas secundarias con el calificador de rol ^r</b>	Entradas secundarias con calificadores (720\$a, \$e; 700\$a, \$e; 710\$a, \$e)	<b>6. Contribuidores:</b> personas u organismos que, aparte del CREADOR, han hecho contribuciones intelectuales importantes al recurso, tales como editores, transcritores, ilustradores, comentaristas, o diseñadores de homepages.
62. <b>Publicador o editor.</b>	Publicador, distribuidor, pie de imprenta (260\$b)	<b>5. Publicador:</b> el organismo responsable de poner el recurso a la disposición en su forma presente (publicador, departamento universitario, organismo internacional, etc.). La idea es indicar a dónde se puede obtener el recurso.
64 y 65. Fecha y fecha normalizada.	<b>Fecha de publicación, distribución, etc. (260\$c).</b> Cuando se modifica el registro se usa el 005 para la fecha de la última corrección.	<b>7. Fecha:</b> la fecha en que el recurso se puso a la disposición, en su forma actual. Solo se usa la fecha normalizada (año, mes y día), 19970320 para el 20 de marzo de 1997.
70.	Entrada no específica (787\$n)	<b>13. Relación:</b> relaciones existentes entre diversos objetos de un documento, imágenes, capítulos de libros o partes de colecciones, series, etc.
73. 74. Cobertura temporal. 80.81.82 Cobertura espacial.	Nota general (500\$a). Para material cartográfico (225\$c). Período cubierto (513\$b).	<b>14. Cobertura:</b> las localizaciones espaciales y duraciones de los recursos.
<b>83. Resumen</b>	Resumen, contenido, etc. (520\$)	<b>4. Descripción:</b> una descripción textual del contenido del recurso, que incluye compendios en el caso de objetos tipo texto, y descripción del contenido en el caso de objetos visuales.
87.88.89. Descriptores principales, secundarios, y no controlados.	Término de indizado no controlado (653\$a). Cuando se usan los encabezamientos LCSH (650\$a). Para la clasificación LC, entonces (050\$a), y si es Dewey, (0802\$a)	<b>3. Tema:</b> descriptores y palabras claves. El tópico del recurso, las palabras claves o frases que describan el tema o contenido. La idea es emplear los tesauros, e incluir aquí las clasificaciones que usan las bibliotecas.