

ARTÍCULO

LA INTEROPERABILIDAD Y EL INTERCAMBIO DE METADATOS EN LA RED

Miguel Ángel G Mejía Argueta

La interoperabilidad y el intercambio de metadatos en la red

Resumen

En el mundo actual, y debido a la generación de una gran cantidad de documentos en Internet, surge la necesidad de organizar mejor esta información para facilitar a los usuarios su búsqueda y localización.. A raíz de esta situación, nacen los metadatos y con ello toda una nueva cultura para la organización de la información, que trae grandes ventajas y también algunas complicaciones.

Los sistemas de organización de la información deben reaccionar ante estos nuevos retos y garantizar el acceso y la interoperabilidad de los datos para lograr un eficiente intercambio de información.

Palabras clave: Metadatos, repositorios, estándares, Dublin Core, ID3, RDF, Darwin Core, protocolos, OAI-PMH, Interoperabilidad.

Interoperability and metadata exchange on the Web

Abstract

In today's world and the daily generation of large number of documents in the Internet, arise the need to be organized this information in order to find more quickly what users search accurately and appropriately. Following this need arise metadata and thus a whole new culture for the organization of information, bringing great advantages and also some complications.

The organization of information systems must react to these new challenges and ensure access and interoperability of data to achieve an efficient exchange of information.

Key words: Metadata, repository, standars, Dublin Core, ID3, RDF, Darwin Core, protocols, OAI-OMH, Interoperability.

Introducción

Con el crecimiento acelerado de las tecnologías de la información, el uso de Internet fue tomando un auge impresionante, con ello creció la cantidad de información disponible para todo tipo de

usuarios: desde el estudiante a nivel primaria, hasta los investigadores de todas partes del mundo y, claro, de la sociedad en general. Esto resultó ser un problema debido a que la manera de acceder a todos estos datos es una tarea muy difícil. Es por ello que se comenzaron a utilizar diversas herramientas para organizar la información. Fue así como surgieron los conceptos: "metadatos", "repositorios" y "sistemas de gestión de contenidos" que facilitan la tarea de organizar toda esta nueva forma de publicar información para que los usuarios puedan localizar datos de interés de la forma más fácil, intuitiva y eficiente.

Los metadatos

Podemos definir a los metadatos como datos altamente estructurados que describen la información, el contenido, la calidad, la condición y otras características de los datos, o como se dice de manera más simple: "datos sobre los datos".¹

En el caso de los metadatos existen una gran diversidad debido a:

1.- La existencia de diferentes tipos de usuarios.

Usuarios que requieren términos para acceder a documentos.

Investigadores que requerirán amplia información de un recurso.

2.- La existencia de diferentes tipos de recursos.

Recursos con una existencia fugaz.

Recursos con un valor comercial.

Recursos de gran valor para la investigación.

1 Que son los metadatos?" <http://antares.inegi.org.mx/metadatos/metadat1.htm> [Visitado el 11 de Junio de 2011]

3.- La existencia de diferentes proveedores de información.

Servicios de información comercial.

Servicios de información académica y para la investigación ²

Los metadatos aportan valor semántico a los datos incluidos dentro de una página web, esto quiere decir que los buscadores pueden encontrar información asociada al código de las páginas.³

Principales razones para usar metadatos

Los beneficios derivados de la utilización de metadatos son diversos y dependen del área en que se utilicen. En términos generales:

A los creadores de conjuntos de datos, los metadatos les ayudan a:

- evitar duplicaciones de información
- identificar cada conjunto de datos sin ambigüedades
- distribuir información con garantías
- difundir las características de los datos producidos
- reducir carga de trabajo
- disponer de un inventario de la información que se produce.

A los usuarios de los datos, los metadatos les ayudan a:

- encontrar los datos buscados
- conocer información que es clave en los datos
- comprender en profundidad la información
- localizar datos (dentro y fuera de la organización)
- transferir e interpretar los datos correctamente.

² Martínez Arellano, Filiberto Felipe, “Diversidad de los metadatos”, Colegio de Bibliotecología en la Facultad de Filosofía y Letras. UNAM, <http://cuib.unam.mx/~felipe/metadata2000/diversidad.htm>. [Visitado el 01 de Junio de 2011]

³ Metadatos y recuperación de información” <http://sistemasavanzadosderecuperaciondeinformacion.iespana.es/> [Visitado el 01 de Junio de 2011]

A las organizaciones, los metadatos les ayudan a:

- proteger sus inversiones
- limitar sus responsabilidades
- crear la memoria institucional
- compartir datos con otras organizaciones
- reducir costos
- ganar tiempo y dinero
- proporcionar información sobre fuentes y calidades
- mejorar los procedimientos de gestión de datos.

Los metadatos nos ayudan a ubicar de una manera más eficiente a los datos.

La metodología para implementar los metadatos, aunque no es muy simple, se puede resolver de la siguiente manera:

- El primer paso incluye la recopilación, evaluación, procesamiento y almacenamiento de volúmenes de datos.
- Como segundo paso, se tendrá que analizar esta información almacenada y determinar cuál es el proceso de clasificación natural de los datos
- En el tercer paso, conviene revisar si ya existe un estándar para el tipo de datos que se va a clasificar, si este fuera el caso:
 - o Se analizan los diferentes estándares y se aplicará el que más se adapte a las necesidades de descripción de los datos.
 - o En caso de no adaptarse ningún estándar de metadatos para la aplicación a los datos, se deberá de generar un equipo de trabajo para realizar un estándar propio.
- Realizar la clasificación de los datos, ya sea con el estándar propio o el estándar adaptado.

Para definir metadatos de calidad es recomendable realizar las siguientes actividades:

1. Formar un equipo de desarrollo para establecer metadatos.
2. Capacitar al equipo de desarrollo de los metadatos.
3. Definir el propósito del equipo de desarrollo de los metadatos.
4. Tomar en cuenta a los usuarios, ya que ellos saben cómo se clasifican los datos
5. Investigar campos potenciales de metadatos
6. Definir criterios de metadatos, tales como:
 - Nombre del Campo.
 - Tipo de campo (texto, fecha, lista desplegable, numérico, etcétera.)
 - Formato.
 - Límite de caracteres.
 - Valores.

Estándares de metadatos utilizados en la red

Existen dos ramas para objetos digitales que impulsan el desarrollo de estándares en los metadatos, estos son: la multimedia y la web semántica.

La rama de la multimedia proporciona estándares para la descripción de recursos multimedios (audio, video e imágenes).

En cambio, la rama de la web semántica proporciona estándares para la descripción de cada objeto digital (incluyendo los multimedios), así como la interconexión de estos objetos.

Algunos ejemplos de estándares internacionales de metadatos son:

Dublin Core

Darwin Core

ID3

Intercambio de interoperabilidad en los metadatos

La interoperabilidad ha sido una necesidad natural hacia los metadatos y se refiere a la facilidad de las herramientas informáticas para compartir, recuperar y hacer uso de los metadatos.

Actualmente en el mundo se están llevando a cabo diversos esfuerzos para establecer estándares para esquemas de metadatos, pero cada uno se centra en sus propios requerimientos a la hora de definir los atributos, su significado, los posibles valores que pueden tomar y la relación entre ellos. Desde el punto de vista del usuario, que pretende buscar recursos en distintos entornos, esto constituye un problema real, porque le obliga a disponer de las herramientas apropiadas para cada caso, o ser consciente de las diferencias que existen entre ellos.

Actualmente en las bibliotecas digitales, hemerotecas digitales, centros de información, etcétera, se han preocupado por establecer herramientas para el correcto manejo de los metadatos, la recuperación de los mismos y de la información a la que hacen referencia. Esto ha traído nuevos retos para la informática.

En este esquema, la UNAM no se ha quedado atrás y en la actualidad cuenta con varios proyectos de repositorios institucionales donde se aplica el intercambio de metadatos.

En estos proyectos han colaborado diversas instituciones de la UNAM y por la Dirección General de Cómputo y Tecnologías de la Información (DGCTIC), concretamente, el área de Acervos Digitales.

Estos son:

Red de Acervos Digitales (RAD) donde existe un intercambio de metadatos, cuyo sitio es : <http://www.rad.unam.mx/index.php/index>

Y el proyecto llamado Latindex Portal de Portales: <http://www.latindex.ppl.unam.mx/>

En ambos casos se ha utilizado el protocolo OAI-PMH para compartir, intercambiar y cosechar metadatos.

Protocolo OAI-PMH (Open Archives Initiative-Protocol Metadata Harvesting)

El protocolo OAI-PMH (Protocolo OAI para la Recolección de Metadatos) es una herramienta de interoperabilidad, independiente de la aplicación, que permite realizar el intercambio de información para que desde puntos centralizados (proveedores de servicio) se puedan realizar búsquedas conjuntas sobre los metadatos de todos aquellos repositorios asociados (proveedores de datos) que, de otra manera, no sería posible simplemente por desconocimiento de su existencia.

El protocolo OAI-PMH en el caso de recuperación de datos funciona de la siguiente manera:

El usuario realiza una petición que es enviada al proveedor de servicios, el cual, por medio del protocolo OAI-PMH, usa transacciones HTTP e idea un proveedor de datos que le regrese los metadatos almacenados en los recursos localizados dentro de un repositorio o contenedor. Estos metadatos son enviados primero al proveedor de datos, quien a su vez los envía al proveedor de servicio, en donde se encuentran instalados los Harvest o recolectores. Dichos recolectores procesan los metadatos, que arrojarán todos los resultados de la búsqueda en formato xml, para que el usuario, al acceder a los recursos mostrados, pueda ver los documentos que se encuentran en el repositorio.

La iniciativa OAI nace del movimiento de Eprints cuyo objetivo es poner a disposición del público documentos en formato electrónico vía repositorios de documentos. En la actualidad existen una serie de programas que facilitan esta tarea y permiten a cualquier institución (universidad o centro de investigación) crear su propio archivo compatible con OAI-PMH. Algunos ejemplos de estos programas son:

Eprints.- Es un software desarrollado por Stevan Harnad, en el Departamento de Ciencias de la Computación y Electrónica de la Universidad de Southampton, en el Reino Unido, con la colaboración de Mike Jewell. El diseño y desarrollo de la versión 1.0 data del año 2001, estuvo a cargo de Robert Tansley -basado en el software CogPrints, desarrollado por Matt Hemus. Se distribuye bajo la licencia de GNU. Es muy fácil de instalar y gratuito, es por ello que ha tomado gran popularidad.

Dspace.- Es un software de código abierto que permite el intercambio libre de colecciones Digitales. Fue liberado en el 2002, como producto de una alianza de HP y el MIT, bajo una licencia BSD y es totalmente compatible con el protocolo OAI-PMH. Puede correr sobre Base de Datos Postgres u Oracle.

CDSware. Su primera versión se hizo pública en Agosto de 2002. Está desarrollado, mantenido y utilizado por el CERN Document Server de Ginebra. Su objetivo es permitir a una institución crear su propio servidor de eprints, catálogos de sus fondos o un sistema documental a través del web. Es compatible con OAI-PMH. Lo más destacado es que utiliza el formato MARC 21 para almacenar los registros bibliográficos. Al igual que en los casos anteriores es un software gratuito distribuido bajo la licencia GPL (General Public License).

Dublin Core y OAI-PMH

Como se puede notar, Dublin Core es una manera muy sencilla y práctica de determinar los metadatos para la búsqueda de diversos recursos. Desde sus inicios ha sido bien recibido por la mayoría de los desarrolladores de la industria informática.

Cuando hablamos de intercambio de recursos digitales, la manera, hasta ahora, más eficiente de acceder a ellos es por medio de los metadatos, debido a que estos proporcionan una descripción de dichos recursos. En el caso de Dublin Core, ha brindado una forma mucho más ordenada y detallada de describir a los documentos. Bajo este sistema, la estandarización de los metadatos permitiría que la indización y acceso a estos recursos sea más ágil y si pensamos en el futuro, se pueden crear nuevas herramientas para clasificar, buscar e identificar todo tipo de documentos.

El caso del uso del protocolo OAI-PMH es de suma importancia, ya que este es un protocolo especializado únicamente en la transferencia de recursos digitales, principalmente de carácter científico y de acceso libre, que puede realizar de un repositorio a otro. Han existido otros tipos de protocolos que tienen el mismo objetivo., la diferencia entre ellos es que OAI-PMH es más simple, ya que se basa en la comunicación entre repositorios utilizando metadatos sujetos a criterios de búsqueda, lo cual hace mucho más eficiente el trabajo.

Tener un estándar que pueda adaptarse sumamente bien a este protocolo hace que sea mucho más fácil y sencilla la transferencia de dichos recursos.

Dublin Core y OAI-PMH

1. En el mundo actual, donde día a día hay cada vez mayor acceso a la información y se gesta una gran cantidad de documentos disponibles en la Internet, es necesario contar con mecanismos de recuperación eficientes y adecuados.
2. Los metadatos surgen como una forma para describir, de manera detallada, los documentos digitales y por ende, nace la necesidad de organizar y controlar también estos metadatos.
3. Los estándares para el manejo de metadatos se originan a partir de necesidades concretas, es decir: Darwin Core para colecciones naturales, Dublin Core y RDF para bibliotecas digitales, ID3 para objetos multimedia, etcétera.
4. La interoperabilidad actualmente está en una primera etapa y es relativa, debido a que están surgiendo estándares de metadatos sujetos al materia de trabajo y en realidad no existe un modelo homogéneo para todos los diferentes tipos de objetos digitales.

5. Al intercambio de metadatos entre diversas herramientas informáticas se le denomina cosecha de metadatos y al software informático que lo puede realizar, de manera automática o semiautomática, se le llama cosechador (harvester).

6. Estos cosechadores permiten no sólo la cosecha, sino el intercambio de los metadatos y realizar búsquedas muy eficientes por la red para cosechar más metadatos. Esto permite a un cosechador obtener muchos metadatos y encontrar el objeto dueño de sus metadatos con mayor precisión en una búsqueda determinada.

Bibliografía

Sistema de Información Geoespacial”, http://www.atlasdebuenosaires.gov.ar/mapas_pdf/Descripcion_Metadatos_AABA.pdf [Visitado el 1 de Junio de 2011]

Perez Velandia, Mayerly y Luis Felipe Silva, “Como Funciona el protocolo OAI-PMH en la recuperación de la información”, <http://www.scribd.com/doc/13073011/Como-Funciona-El-Protocolo-Oai-Pmh-en-La-Recuperacion> [Consultado 3 de Junio de 2011]

Barrueco, José Manuel, “OAI-PMH: Protocolo para la transmisión de contenidos en internet”. www.uv.es/barrueco/cardedeu.doc [Visitado el 3 de Junio de 2011]

“The Open Archives Initiative Protocol for Metadata Harvesting” <http://www.openarchives.org/OAI/openarchivesprotocol.html#DefinitionsConcepts> [Visitado el 3 de Junio de 2011]

“OAI para principiantes” <http://travesia.mcu.es/portaln/jspui/html/10421/1823/page4.htm>
[Visitado el 3 de junio de 2011]

Guajardo Salinas, Aldo, “Z39.50 y OAI-PMH: Protocolos de Transferencia y Recuperación de Información”, http://www.bibliotecarios.cl/descargas/2010/11/guajardo_z3950.pdf, [Visitado el 3 de Junio de 2011]

Ferreas Hernandez, Transito, “Acceso Abierto a la comunicación científica: Los repositorios institucionales”, Universidad de salamanca, <http://www.slideshare.net/Transito09/comunicacion-cientifica5> [Visitado 3 de Junio de 2011]

Enric Peig Olivé, “Interoperabilidad de metadatos en sistemas distribuidos”, Universitat Pompeu Fabra, <http://www.tesisenred.net/bitstream/handle/10803/7531/tepo1de1.pdf?sequence=1>
[Visitado 10 de Junio de 2011]