Perfiles semánticos de metadatos. Un método para la vinculación de recursos de información documental en el entorno digital

Eder Ávila Barrientos
Universidad Nacional Autónoma de México

Introducción

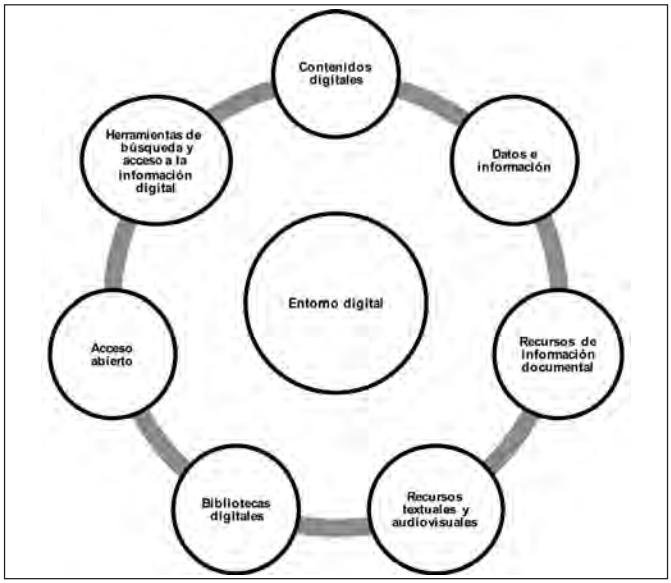
La web semántica es uno de los desarrollos digitales que más incidencia ha tenido en el uso y la generación de grandes cantidades de datos. Asimismo, ha brindado la posibilidad de intercambiar significativas cantidades de información mediante el desarrollo de aplicaciones automatizadas como algoritmos, software especializado y generadores informáticos.

A su vez, las bibliotecas digitales han experimentado cambios notables en sus estructuras que van desde la incorporación de recursos de información digital en sus colecciones, hasta la posibilidad de ofrecer servicios digitales de información con un valor agregado. Estos servicios son destinados especialmente a atender las demandas de las comunidades académicas y de investigación; sin embargo, su alcance debe ser aprovechado para ofrecer servicios a diferentes tipos de comunidades de usuarios.

El entorno digital es un escenario complejo de interacción entre diferentes fuentes de datos e información disponibles en la web. Los datos y la información en este contexto son procesados por sistemas computarizados y altamente especializados. Los datos y

la información generada en el entorno digital es procesada, comunicada y distribuida por medio de las tecnologías de la información y comunicación.

Ilustración 1. Elementos del entorno digital



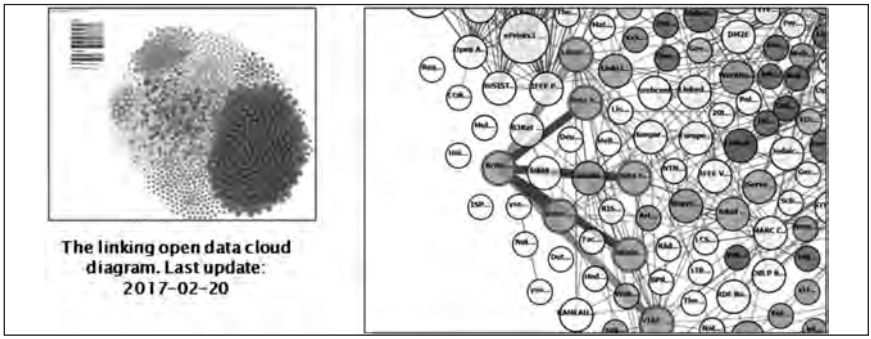
Fuente: elaboración propia.

En el entorno digital, hay una serie de elementos característicos (véase la ilustración 1) que hacen posible la generación de productos y servicios de índole digital, los cuales utilizan datos e información que son generados en diferentes contextos profesionales, académicos, gubernamentales y culturales. Por lo tanto, se tiene la posibilidad de contar con datos e información de diferente temática y naturaleza.

En el ámbito de las bibliotecas digitales, las principales fuentes de datos e información son los catálogos en línea, repositorios institucionales, bibliotecas digitales, bases de datos, repositorios de datos y portales temáticos. Éstos contienen registros bibliográficos

y de autoridad que remiten a recursos de información documental como libros digitales, tesis, recursos audiovisuales, artículos de investigación, imágenes y recursos susceptibles de contener información de índole documental.

Ilustración 2. The linking open data cloud diagram.



Fuente: Abele 2017.

Las bibliotecas digitales y la web semántica han encontrado un escenario de interacción que es ejemplificado en el diagrama de la nube de datos enlazados abiertos (véase ilustración 2). En este diagrama, se puede observar la participación de las bibliotecas mediante la generación de vocabularios controlados que tienen como propósito vincular temáticamente los datos bibliográficos y de autoridad pertenecientes a registros de recursos de información documental.

Un proyecto bastante significativo que resume la importancia de la interacción entre las bibliotecas digitales y la web semántica es el Library Linked Data Incubator Group (World Wide Web Consortium 2011), el cual busca conformar un grupo de especialistas para aumentar la interoperabilidad global de los datos bibliotecarios en la web por medio de los datos vinculados que son generados por la comunidad bibliotecaria, así como desarrollar iniciativas e identificar oportunidades de colaboración para el futuro.

Como parte del reporte final del grupo, se menciona la posibilidad de establecer vinculaciones entre datos bibliográficos y de autoridades disponibles en las bibliotecas, tomando en cuenta específicamente el procesamiento de los datos que están disponibles en las herramientas para la búsqueda y el acceso a la información digital. Partiendo de esta posibilidad, este trabajo tiene como objetivo determinar la función de los perfiles semánticos de metadatos como un método para la vinculación de recursos de información documental en el entorno digital. Para ello, se analiza la presencia de la web semántica y su relación con los contenidos disponibles en el entorno digital; se estudian los atributos de los perfiles semánticos de metadatos, y se examina el proceso de vinculación de recursos de información documental en el entorno digital.

La web semántica y la relación de contenidos en el entorno digital

El concepto de web semántica pone de manifiesto una serie de elementos técnicos, pragmáticos y teóricos que caracterizan a un entorno digital con significado. Los datos que la conforman son de diversa índole; por ejemplo, geográficos, gubernamentales, administrativos y, por supuesto, bibliográficos o documentales.

> La comprensión crítica de la web semántica es particularmente importante para los profesionales de la información que contribuyen y hacen uso de la web como un gran ambiente para ofrecer servicios de información. Se argumenta que la comprensión crítica de los diferentes discursos asociados con la web semántica resultarán en pensar, optimizar y mejorar la generación y presentación de los diferentes servicios y productos de información (Calaresu y Shiri 2015, 97).

Las herramientas de la web semántica posibilitan la generación de servicios digitales de apoyo a la investigación científica y académica. Los servicios planteados por Calaresu y Shiri serán de valor agregado ya que los datos que los conformarán pasarán por procesos

intelectuales, analíticos y técnicos gestionados por los profesionales de la información.

Los datos bibliográficos forman parte de registros catalográficos, descriptivos y representativos que están disponibles en herramientas de acceso a la información documental como catálogos en línea, bases de datos, repositorios digitales, descubridores de información y portales temáticos.

Las bibliotecas digitales cobran relevancia porque son la fuente principal de los datos bibliográficos. También proporcionan acceso a datos documentales que representan una determinada colección de recursos de información documental, y a datos de autoridad que representan a los creadores de las obras.

En el ambiente digital, hay una amplia gama de contenidos reunida, ésta va desde la textual hasta los sofisticados contenidos audiovisuales. Los principios de la web semántica crean relaciones que posibilitan la unión de contenidos con significados similares; por ejemplo, relaciones entre autores, temas, expresiones y manifestaciones de una determinada obra. El establecimiento de estas relaciones se gesta a partir de un proceso intelectual en donde intervienen vocabularios especializados y una normatividad técnica que hace posible la creación de entornos de relaciones semánticas, como la Resource Description Framework (RDF) y la Ontology Web Language (OWL).

La integración de RDF y OWL en un entorno específico de datos posibilita el establecimiento de relaciones significativas entre ellos. Las relaciones de significado entre los datos disponibles siempre tendrán una relación directa con la necesidad del usuario final, ya que el significado es expresado por dicho usuario al momento de demandar información.

La relación de contenidos en el entorno digital pone de manifiesto un proceso que pretende la identificación del universo documental que rodea a los recursos de información mediante el establecimiento de vinculaciones significativas para el usuario final de acuerdo con su contexto y demanda informativa. Esta singularidad otorga la posibilidad de establecer vinculaciones entre recursos disponibles en diferentes contextos de la web además de

propiciar la interacción de la biblioteca con entornos de índole digital mediante la expansión de sus servicios y colecciones.

Perfiles semánticos de metadatos

Las estructuras descriptivas de metadatos que permiten representar y vincular los atributos bibliográficos y de autoridad de los recursos de información documental son conocidas como perfiles semánticos de metadatos. Éstos se construyen a partir de un análisis intelectual del recurso y el contexto en donde están inmersos, ya sea una colección en específico, en un entorno digital o dirigidos a atender las demandas de una comunidad de usuarios en particular.

Estos perfiles están conformados por una estructura semántica que hace posible la vinculación de los atributos mediante el registro de datos. En este sentido, la vinculación semántica le otorga un significado al recurso; es decir, se establece un proceso que va más allá de simplemente cubrir una demanda informativa del usuario final, se trata de un método para descubrir el universo documental que rodea al recurso. Así pues,

> [...] la semántica es un término usado tanto por lingüistas para describir el estudio de los significados. En un nivel muy básico, el campo explora cómo se codifican los conceptos en un lenguaje a través de un sistema de símbolos que facilita la comunicación entre entidades (Calaresu y Shiri 2015, 83).

Por lo tanto, el significado de una vinculación tendrá diferentes interpretaciones dependiendo del contexto en donde ésta se desarrolle y de la necesidad informativa del usuario. El proceso para construir perfiles semánticos de metadatos consta de las siguientes etapas:

- Identificación de atributos del recurso. Los atributos de los recursos de información digital son únicos e independientes. Éstos se representan en registros de índole bibliográfico y de autoridad. Para ello, se utilizan normas de carácter

bibliotecario como lenguajes documentales y principios de catalogación. La identificación de atributos de un recurso es un proceso minucioso que involucra la ubicación de elementos descriptivos.

- Descripción de los atributos. Los elementos descriptivos identificados son utilizados para representar el contenido intelectual, bibliográfico y creativo de los recursos de información documental. Este proceso conlleva el uso de esquemas de metadatos y lenguajes documentales como tesauros y ontologías para establecer descripciones precisas y significativas para el usuario final.
- Normalización de los datos. La uniformidad y normalización de los datos es necesaria para contar con datos consistentes y capaces de vincularse con datos similares que se encuentran en su mismo entorno o en fuentes externas a los mismos.
- Establecimiento de vinculaciones. Los identificadores uniformes de recursos (URI) son el mecanismo indispensable para establecer las vinculaciones entre los datos resultantes. Cada dato como título, autor y tema deberá contar con un URI independiente y univoco. La asignación de URI a los datos es un procesamiento sistemático que se desarrolla con el uso de aplicaciones informáticas y la gestión de enlaces proporcionada por un sistema de información.
- Mecanismo de visualización. La generación de datos vinculados requiere de una interfaz de visualización que permita al usuario final buscar, recuperar, acceder y visualizar los datos y sus respectivas vinculaciones. Para ello, se utilizan normas informáticas y lenguajes de programación. El más utilizado para este propósito es sparql.
- Acceso abierto a la vinculación. Todo proyecto de datos vinculados requiere de una política de acceso abierto que permita al usuario final la libre consulta de los datos sin restricciones técnicas, económicas y legales. El verdadero potencial de los datos vinculados está en su capacidad de establecer relaciones con fuentes externas a su contexto.

Para que esto sea posible, es necesario contar con un marco de políticas que contemple la vinculación y el acceso abierto a los datos.

Ilustración 3. Sintaxis XML/RDF de un perfil semántico de metadatos correspondiente a una monografía digital

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/terms/">
  <rdf:Description rdf:about="http://bit.ly/2rb0eHX">
    <dc:title>Biblioteca digital académica en bibliotecología y estudios de la información</dc:title>
    <dc:creator> Torres Vargas, Georgina Araceli</dc:creator>
    <dc:subject>Bibliotecas digitales académicas</dc:subject>
    <dc:subject>Recursos de información electrónicos</dc:subject>
    <dc:subject>Tecnología de la información</dc:subject>
     <dc:publisher>UNAM</dc:publisher>
     <dc:publisher>Instituto de Investigaciones Bibliotecológicas y de la Información</dc:publisher>
     <dc:contributor>Georgina Araceli Torres Vargas</dc:contributor>
     <dc:contributor> Ariel Alejandro Rodríguez García</dc:contributor>
     <dc:contributor>Ana Lúcia Terra</dc:contributor>
     <dc:contributor>Miguel C. Gama Ramírez</dc:contributor>
     <dc:contributor>Adriana Monroy Moñorz</dc:contributor>
     <dc:contributor>Francisco José Valentín Ruiz</dc:contributor>
     <dc:contributor>Juan Miguel Sánchez Vigil</dc:contributor>
     <dc:contributor>Edel Ávila Barrientos</dc:contributor>
     <dcterms:date:created>2013</dcterms:date:created>
     <dc:type xsi:type="dcterms:DCMIType">Monografía de investigación</dc:type>
     <dc:type xsi:type="dcterms:DCMIType">Text</dc:type>
     <dc:format>Digital</dc:format>
     <dc:format xsi:type="dcterms:IMT">application/pdf</dc:format>
     <dc:identifier>"http://bit.ly/2rb0eHX"</dc:identifier>
     <dc:language xsi:type="dcterms:ISO639-2">spa</dc:language>
     <dc:language xsi:type="dcterms:ISO639-2">por</dc:language>
     <dc:relation:is part of>Biblioteca digital en bibliotecología y estudios de la información</dc:relation:is part of>
     <dc:relation:is part of>Políticas de informacao para bibliotecas digitais: conceitos e componentes</dc:relation:is part of>
     <dc:relation:is part of>El desarrollo de colecciones de libros electrónicos en las instituciones de educación superior</dc:relation: is part of>
     <dc:relation:is part of>Fotografía digital: catalogación y metadatos para la biblioteca digital</dc:relation: is part of>
     <dc:relation: is part of>Del libro electrónico a las bibliotecas digitales en España: estado de la cuestión</dc:relation:is part of>
     <dc:relation:is part of>Linked Open Data en la biblioteca digital semántica académica</dc:relation:is part of>
     <dc:rights>http://creativecommons.org/licenses/by-nc-sa/3.0/deed.es_MX</dc:rights>
      <dcterms:educationlevel>Nivel superior</dcterms:educationlevel>
  </rdf:Description>
</rdf:RDF>
```

Fuente: elaboración propia.

En la ilustración 3, puede apreciarse la conformación de un perfil semántico de metadatos correspondiente a una monografía digital que es representada mediante XML/RDF. La semántica del perfil se encuentra definida a partir de los elementos descriptivos utilizados para representar el contenido intelectual, bibliográfico y creativo del recurso. En este sentido, "la creación fluida de metadatos RDF depende de dos cosas: una ontología predefinida que le ahorre al autor la tarea de crear términos y relaciones y una interfaz fácil de usar que permita al autor crear instancias de metadatos intuitivamente" (Greenberg 2003, 18). Es importante mencionar que la semántica de los datos vinculados es generada a partir de la demandas de los usuarios finales.

Una necesidad informativa puede contener un significado en específico para una comunidad en particular o para un grupo específico de usuarios finales. La función principal de los datos vinculados es manifestar y representar el significado de las demandas de información mediante la relación de datos con atributos similares y altamente relevantes para el usuario final.

Vinculación de recursos de información documental en el entorno digital

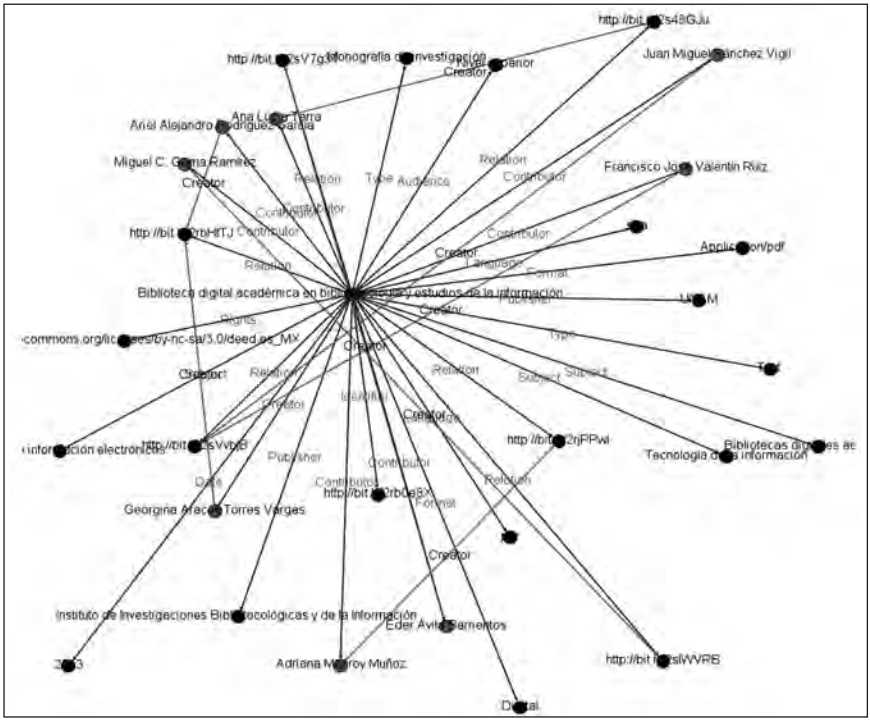
El establecimiento de vinculaciones de recursos de información documental en el entorno digital utiliza perfiles semánticos de metadatos definidos para cada tipo de recuso en particular. En el caso de las monografías digitales, los elementos descriptivos que representan los atributos del recurso han sido seleccionados a partir de la utilización del esquema de metadatos Dublin Core. Se ha utilizado la norma RDF para definir las vinculaciones semánticas de los datos, el lenguaje XML para la interconexión web de los datos y la asignación de URI para establecer las relaciones existentes entre los datos disponibles.

En la ilustración 4, vemos una representación gráfica de la función del perfil semántico de metadatos correspondiente a una monografía digital. Se ha utilizado la aplicación informática Gephi para llevar a cabo el procesamiento de los datos y obtener una visualización de las relaciones obtenidas mediante un grafo RDF. Esta representación de los datos permite consultar las relaciones existentes entre los datos bibliográficos y de autoridad correspondientes a un registro que representa los atributos de una monografía digital.

Mediante el título de la monografía, es posible identificar las relaciones existentes entre sus colaboradores, los temas investigados en el contenido intelectual de la obra y las diferentes vinculaciones entre los datos bibliográficos del recurso. Por lo tanto, la vinculación de recursos de información documental en el entorno digital requiere de una interoperabilidad entre las fuentes que

serán vinculadas mediante los perfiles semánticos de metadatos. Es decir, los contextos digitales que intenten vincular sus respectivos recursos deberán utilizar las mismas normas, principios y lenguajes para su implementación.

Ilustración 4. Función del perfil semántico de metadatos relativo a una monografía digital



Fuente: elaboración propia.

En un entorno semántico de datos, las vinculaciones establecidas son de gran utilidad para comprender el movimiento de un determinado fenómeno. En el caso de los recursos de información documental, los datos vinculados ayudan a comprender la manera en la que se comporta la producción intelectual de un determinado

campo de conocimiento, pues cualquier recurso arbitrado y validado es producto del proceso de una investigación científica, académica y profesional.

Por lo tanto, el acceso a las vinculaciones entre los datos disponibles en un contexto digital otorga al usuario final la posibilidad de identificar el universo documental que rodea a un recurso, su autor y su tema, o un tipo de expresión y manifestación realizada como parte de un proceso intelectual.

En el 2017, la International Federation of Library Associations and Institutions (IFLA) publicó el modelo de referencia para bibliotecas: un modelo conceptual para información bibliográfica que tiene como objetivo establecer principios generales explícitos que rijan la estructura lógica de la información bibliográfica sin hacer presuposiciones sobre cómo esos datos pueden almacenarse en un sistema o una aplicación en particular. Mediante la estructura lógica de estos datos, se plantea la posibilidad de ejercer una referenciación de los mismos a través el seguimiento de un modelo entidad-relación que respalde la propuesta. Es decir, establecer relaciones entre los datos disponibles en registros bibliográficos que remiten a recursos de información documental disponibles en las colecciones de bibliotecas.

En este sentido, cobra vital importancia la generación de mecanismos de visualización de datos vinculados que sean usables, interoperables y accesibles para el usuario final, pues la consulta remota de los datos es un aspecto que involucra el acceso eficiente a éstos, mediante la generación de interfaces eficientes que cumplan con este propósito. Los principios de usabilidad, interoperabilidad y accesibilidad proporcionan los elementos para la correcta arquitectura de las interfaces de usuario. Estos principios pueden tomarse en cuenta para generar las aplicaciones que permitan la interacción de los datos vinculados con los usuarios finales. Es decir, que otorguen la posibilidad de buscar, recuperar, acceder y visualizar las relaciones entre los datos existentes.

Los perfiles semánticos de metadatos pueden utilizarse como un método para descubrir el universo documental que rodea a un dato en específico sin la necesidad de efectuar búsquedas avanzadas

y con el potencial de simplificar el descubrimiento de recursos de información documental y sus respectivas vinculaciones semánticas con otras fuentes de datos disponibles en la web.

Los esquemas de metadatos juegan un papel trascendental en el desarrollo de aplicaciones que hagan posible la construcción de entornos digitales de datos semánticos por medio de su integración con principios normativos como los proporcionados por las ontologías y los lenguajes documentales. Open Metadata Registry es una plataforma web que apoya la generación de perfiles semánticos de metadatos interoperables con esquemas que han sido adoptados para su uso en las bibliotecas:

> El registro proporciona un medio para identificar, declarar y publicar esquemas de metadatos (conjuntos de elementos/propiedades), esquemas (vocabularios controlados) y perfiles de aplicaciones (AP). Por lo tanto, el registro apoya los objetivos clave de descubrimiento, reutilización, estandarización e interoperabilidad de metadatos a nivel local y global (Open Metadata Registry 2017, s.p.).

Así pues, los perfiles semánticos de metadatos integran en su estructura un aparato normativo que tiene como objetivo conformar una estructura descriptiva y representativa de los datos pertenecientes a un recurso de información documental con la posibilidad de vincular sus atributos de manera significativa para el usuario final.

Las vinculaciones significativas entre los datos serán el resultado del análisis de las demandas informativas de la comunidad de usuarios a la cual estarán dirigidos los datos. Es decir, el significado de los datos enlazados está influenciado por la demanda informativa del usuario final, y se manifiesta normativamente mediante el uso de ontologías y lenguajes documentales. Por lo tanto, la generación de perfiles semánticos de metadatos es un proceso que nace de la demanda informativa del usuario.

Consideraciones finales

La generación de perfiles semánticos de metadatos otorga la posibilidad de vincular significativamente los registros pertenecientes a recursos de información documental, disponibles en el entorno de la web, sobre todo en aquellos espacios en donde se localizan las herramientas de acceso a la información documental como catálogos en línea, bases de datos, repositorios digitales y portales temáticos.

La concentración de estos registros y su vinculación ofrecen la posibilidad de generar servicios de información innovadores con un valor agregado que sean destinados para el apoyo a la investigación y la academia. Tal es el caso de la conformación de repositorios de datos de investigación, una herramienta que permite gestionar, buscar, recuperar y acceder a los conjuntos de datos derivados de la investigación científica y académica.

En este sentido, los Estudios de la Información se enfrentan a un fenómeno de investigación que es necesario abordar desde su óptica, pues no se trata sólo de la generación de nuevos servicios, sino de la aparición de nuevas demandas informativas por parte del usuario final y su interacción con el entorno digital.

Bibliografía

Abele, Andrejs *et al*. The Linking Open Data Cloud Diagram 2017. Disponible el 20 de agosto de 2017 en http://lod-cloud.net/.

Calaresu, Michael y Ali Shiri. "Understanding Semantic Web: a Conceptual Model". *Library Review*, 64 ½, 2915 (2015): 82-100.

Greenberg, Jane, Stuart Sutton y Grant Campbell. "Metadata: A Fundamental Component of the Semantic Web". *Bulletin of the American Society for Information Science and Technology.* April/May (2003): 16-18.

International Federation of Library Associations and Institutions (IFLA). IFLA Library Reference Model: A Conceptual Model for Bibliographic Information. Pat Riva, Patrick Le Boeuf y Maja Žumer, eds. Países Bajos: IFLA, 2017.

Open Metadata Registry. Página web. 2017. Disponible el 25 de agosto de 2017 en http://metadataregistry.org/about.html.

World Wide Web Consortium. Library Linked Data Incubator Group Final Report. W3C Incubator Group Report. 2011. Disponible el 21 de agosto de 2017 en https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/.