

OPEN DATA Y LINKED OPEN DATA: SU IMPACTO EN EL ÁREA DE BIBLIOTECAS Y DOCUMENTACIÓN

Fernanda Peset, Antonia Ferrer-Sapena e Imma Subirats-Coll



Fernanda Peset es licenciada en geografía e historia (1990) y doctora en documentación (2002). Profesionalmente viene del mundo de las unidades de información y desde 1997 es profesora titular de la *Universidad Politécnica de Valencia*. Su trabajo y publicaciones se orientan a la comunicación científica, acceso abierto e implantación del protocolo OAI-PMH, normalización de la información, descripción de documentos, sistemas de documentación de museos, etc. Es responsable del *Posgrado Oficial Industrias Culturales y de la Comunicación*. Participa en proyectos como *IraLIS*, *E-LIS* y el *Grupo Ciepi*.

Universidad Politécnica de Valencia
Camino de Vera, s/n. 46022 Valencia
mpesetm@upv.es



Antonia Ferrer-Sapena, licenciada en geografía e historia en la especialidad de historia contemporánea por la *Universidad de Valencia* y doctora en técnicas y métodos actuales en información y documentación, es profesora titular de la *Universidad Politécnica de Valencia (UPV)* y coordinadora de investigación en *Florida Universitaria*. Ha coordinado numerosos proyectos nacionales e internacionales. De los últimos proyectos dirigidos cabe destacar el portal de economía social *EcSocial* (<http://www.ecsocial.com>) y la *Biblioteca Digital del Cooperativismo Valenciano* (<http://www.focoop.es>).

Depto. de Comunicación Audiovisual, Documentación e Historia del Arte, UPV
Camino de Vera, s/n. 46022 Valencia
anfersa@upv.es



Imma Subirats-Coll es especialista en gestión de la información en la *Food and Agriculture Organization (FAO)* de la *ONU* desde 2006. Antes trabajó en el *Gobierno de Catalunya* en Barcelona. Estudió historia y ciencias de la información en la *Universidad de Barcelona*. Tiene dos masters: en historia, y en biblioteconomía y ciencias de la información. Responsable de la gestión de contenidos del proyecto *Archivo Abierto FAO (FAO OA)*, liderando la integración de las normas MODS y FRBR. Apóstol del acceso abierto, especialmente a través de la red agrícola *Ciard*. Co-fundadora y la coordinadora ejecutiva de *E-LIS (E-prints en Biblioteconomía y Ciencias de la Información)*.

Food and Agriculture Organization (FAO)
Viale delle Terme di Caracalla, 00153 Roma, Italia
imma.subirats@gmail.com

Resumen

Se da cuenta de los primeros pasos de las iniciativas *Open data* y *Linked open data* como generadoras de innovación en el sector de la gestión de la información. Tal y como supusieron los movimientos *Open access (OA)* y *Open archives initiative (OAI)* hace unos años, estas dos iniciativas constituirán un revulsivo que conmocionará la innovación tecnológica y la estructura de la Web. Nuestro trabajo presenta el panorama mundial de *open data*, así como los proyectos nacionales relacionados con los datos de los gobiernos. Cuatro administraciones españolas han comenzado ya a liberar sus datos, por orden cronológico: Asturias, Euzkadi, Cataluña y el Ayuntamiento de Zaragoza. En el caso de *linked open data* se ofrece también el panorama mundial, con 203 proyectos registrados. Describimos con más profundidad el trabajo que el *W3C* está llevando a cabo en el ámbito de las bibliotecas. Por último se presentan las implicaciones que pueden suponer estos nuevos avances para nuestro sector, la documentación.

Palabras clave

Linked open data, Open data, Open government, Library Linked Data Incubator Group, Datos abiertos, Web semántica, Datos enlazados.

Title: *Linked open data and Open data. Its impact in the field of libraries and information science*

Abstract

We show the first steps of the initiatives Open data and Linked open data as sources of innovation in the field of information management. As the Open Access movement (OA) and the Open Archives Initiative (OAI) implied some years ago, these

Artículo recibido el 15-02-11
Aceptación definitiva: 07-03-11

two initiatives constitute a shock that reverberated in technological innovation and the structure of the Web. We present a global overview of open data and national projects related to government data. Four Spanish local or regional administrations have already begun to release their data: in chronological order, Asturias, Euskadi, Zaragoza and Catalonia. In the case of linked open data we also provide a global picture, with 203 registered projects. We describe in more detail the work that the W3C is currently carrying out in the field of libraries. Finally we present the consequences these new developments can have for librarianship and information science.

Keywords

Linked open data, Open data, Open government, *Library Incubator Group Linked Data*, Semantic web.

Peset, Fernanda; Ferrer-Sapena, Antonia; Subirats-Coll, Imma. "Open data y Linked open data: su impacto en el área de bibliotecas y documentación". *El profesional de la información*, 2011, marzo-abril, v. 20, n. 2, pp. 164-172.

DOI: 10.3145/epi.2011.mar.06

Introducción: contexto y definiciones

El término acceso abierto hasta ahora, en el mundo de la información y las bibliotecas, se ha aplicado mayoritariamente a la promoción del acceso a la documentación científica. Pero hoy día está yendo más allá y se habla del acceso abierto a los datos, refiriéndose específicamente a la preparación y exposición de contenidos usables y reutilizables para la web semántica. Los dos conceptos clave son *Open data* (OD) y *Linked data* o *Linked open data* (LOD), traducible al español como "datos abiertos vinculados (o enlazados)". Ambos se derivan de las declaraciones del creador de la web, **Tim Berners-Lee** (2006), asesor del gobierno británico. Y además el segundo se relaciona con la familia de estándares del *W3Consortium* (2010) para la construcción de la web semántica.

OD es un movimiento que promueve la liberación de datos, generalmente no textuales y en formatos reutilizables como CSV (*comma separated values*), procedentes de organizaciones; el segundo se refiere a la exposición de esos u otros datos en la web, descritos en RDF (*resource description framework*) y con indicación de la relación que puede existir entre esos datos y otros. Los datos liberados pueden ser de cualquier temática: geográficos, meteorológicos, científicos, de tráfico..., y fue el gobierno británico quien comenzó a hacerlo con su *Opening up Government* (2010). Además el 30 de septiembre de 2010 creó una licencia para su reutilización: *Open government licence* (2010). Por su parte, la Unión Europea hace años ya había comenzado a interesarse por la reutilización de los datos, como muestra su *Directiva* sobre estos aspectos en el sector público (2003).

Si bien los conceptos son diferentes entre sí, la filosofía OD engloba LOD y puede aportar gran cantidad de contenidos de interés en la web semántica. El movimiento *Open government* se está convirtiendo en una iniciativa ciudadana que aboga por la transparencia de la administración e incluso tiene derivaciones en temas de alcance social como el debate sobre *Wikileaks* o la liberación forzada de datos públicos (**Belanche**, 2009).

En todo caso los resultados de este movimiento dependerán del tipo de datos que se liberen, y también de cómo se analicen o combinen (**Alonso**, 2009). Las posibilidades de innovación que se abren no tienen límites, pero necesitan nuevos proyectos que pongan en valor estos datos así como otras formas de visualización que están por explorar.

El sector de la documentación no se define, sin embargo, por desarrollar productos y servicios. Más bien trabaja con normalidad con protocolos estandarizados para gestionar sus fondos, por lo que a esta comunidad le resultaría fácil participar con sus contenidos en la web semántica a través del uso de *Linked Data* para publicar sus datos. Este concepto tiene una vertiente técnica, relacionada con la descripción de datos estructurados y su publicación en la web para que puedan estar interconectados y ser más útiles (**Bizer; Heath; Berners-Lee**, 2007). Las bibliotecas constituyen un gran potencial de datos de calidad y fiables, en el sentido utilizado por **Hannemann y Kett** (2010) para establecer los diferentes niveles de confianza y persistencia de los datos; pero hasta ahora sólo los han producido y compartido en su propio medio, con poco contacto con otros colectivos. La oportunidad de participar en la web semántica puede abrir las puertas de nuestra comunidad al mundo web... o perder el tren de las tendencias mundiales.

La diferencia entre la web del hipertexto y la web semántica es que mientras una vincula páginas o documentos en html, la segunda enlaza datos estructurados

Principios básicos del LOD

La principal diferencia entre la web del hipertexto y la web semántica es que mientras una vincula páginas o documentos en html, la segunda aboga por ir más allá del concepto documento y enlaza datos estructurados. En 2006 **Berners-Lee** definió cuatro reglas para la publicación de *Linked data*:

1. usar URIs (*uniform resource identifiers*) identificando los recursos de forma unívoca;
2. usar URIs http para que la gente pueda acceder a la información del recurso;
3. ofrecer información sobre los recursos usando RDF;
4. incluir enlaces a otros URIs, facilitando el vínculo entre distintos datos distribuidos en la web.

Estos principios están definidos como reglas, pero en realidad son más bien recomendaciones o buenas prácticas para

el desarrollo de la web semántica. Es posible publicar datos que cumplan sólo los tres primeros principios, pero el hecho de no aplicar el cuarto los convierte en menos visibles y, como consecuencia, menos reutilizables.

RDF es el marco de descripción de recursos para metadatos en la Web elaborado por el W3C (Méndez, 1999). Se basa en la idea de declarar recursos usando la expresión en la forma sujeto-predicado-objeto. Esta expresión es conocida en la terminología RDF como triple o triplete. Un triplete RDF contiene tres componentes, todos con referencia en un URI:

- *sujeto*, una referencia URI, una persona, o un nodo, es el ente al cual nos referimos;
- *predicado* es la propiedad o relación que se desea establecer acerca del sujeto;
- *objeto* es el valor de la propiedad o del otro recurso con el que se establece la relación.

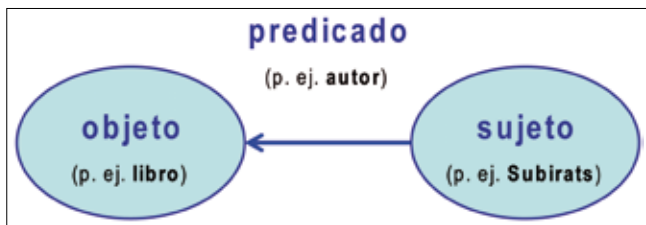


Figura 1. Representación gráfica de un triplete, <http://www.w3.org/TR/rdf-concepts>

El hecho de utilizar URIs para enlazar los datos convierte la web semántica en una especie de gran base de datos que permite que las personas y las máquinas puedan explorar la información referenciada e interconectada entre sí en la Web, lo que al mismo tiempo fomenta su crecimiento.

El uso de LOD permite publicar información migrada desde diferentes fuentes en las que se encuentra almacenada, como por ejemplo bases de datos, ficheros, hojas de cálculo..., siempre usando el lenguaje RDF, para permitir enlazarla con otros datos, de forma que aumente la conectividad entre la información y se facilite así la recuperación y la agregación de información relacionada (Bizer; Cyganiak; Heath, 2007). La web basada en LOD es un avance en la sindicación de contenidos, que utiliza fuentes externas de datos para crear nuevos servicios.

La web basada en LOD es un avance en la sindicación de contenidos, que utiliza fuentes externas de datos para crear nuevos servicios

Situación actual

Los recursos disponibles en *Open data* están experimentando un enorme crecimiento por diferentes sectores (Domínguez, 2010). La delantera en este campo la llevan el Reino Unido y EUA. Para conocer el avance de las fuentes de datos en todo el mundo podemos consultar el proyecto europeo

Public sector information (PSI), que recopila los paquetes de datos que liberan las administraciones.

En España cabe citar las siguientes aplicaciones creadas, según el proyecto *Aporta*:

- mapa de predicciones meteorológicas del *Portal de datos abiertos* de la *Generalitat de Catalunya*;
- mapa de incidencias de tráfico de *Open data Euskadi*, que incluye consejos para la reutilización de los datos y el método seguido;
- callejero del *Portal de Datos Abiertos* de Gijón;
- planificador de visitas del portal de *Datos Abiertos de Zaragoza*, realizada con *Dónde (en) Zaragoza DND* zgz, aplicación desarrollada para móviles.

Algunas acciones de referencia en España son el concurso *Desafío AbreDatos 2010* para creación de APIs (*application programming interface*), los cursos de *MediaLab Prado* de visualización de información, el trabajo de la *Fundación CTIC* y el proyecto *Aporta* financiado por el plan *Avanza* del *Ministerio de Industria, Turismo y Comercio (Micyt)*.

Como acciones de referencia a escala mundial (Mazzo, 2010) encontramos:



Open Knowledge Foundation, iniciativa europea para generar nuevo conocimiento mediante el tratamiento de datos liberados o disponibles en la web. Lidera varios proyectos como *The open definition*, que sienta las bases teóricas; *Data commons* que ofrece soluciones legales; o *CKAN*, que registra los conjuntos de datos y los proyectos de *open knowledge* para facilitar que se encuentren y se usen. *Where does my money go?* es un desarrollo de gran éxito, con información sobre el destino de los impuestos británicos, y *KForge* ofrece las utilidades que son utilizadas en ocasiones para sus propios proyectos construidos sobre conocimiento abierto.



Open Data Foundation, norteamericana, lleva a cabo proyectos con datos estadísticos con la intención de crear estándares válidos para el tratamiento de este tipo de información.



El *W3C* contribuye en este movimiento con un grupo sobre *eGovernment*

En cuanto a *Linked Open Data*, desde 2007 el número de paquetes de datos estructurados se ha incrementado es-

pectacularmente. Las figuras 2 y 3 representan esta evolución (Cyganiak; Jentzsch, 2007-2010). *DBpedia* es una de las principales fuentes de datos que la comunidad usa para aplicar el cuarto principio LOD, y se sitúa como eje fundamental de este proceso.

Con sólo la observación de las figuras 2 y 3, y cuatro años de historia, se constata: 1) la rápida extensión de la iniciativa; 2) la importancia de los proyectos adheridos (por citar alguno, el censo de población norteamericana); 3) la progresiva madurez del sistema, que evoluciona desde la centralidad de una estrella –*DBpedia* en 2007, junto a *FOAF* y *RDF book mashup*– a la explosión de otros proyectos que se convierten en nodos centrales de desarrollos colaterales en 2009,

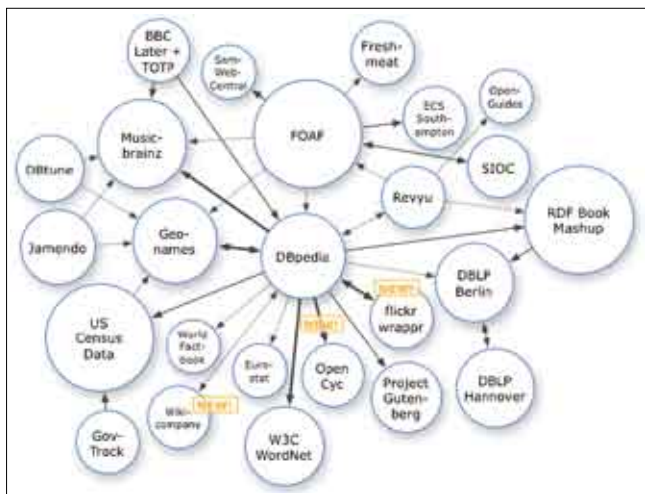


Figura 2. Diagrama de la *Linking data cloud* en 2007 <http://lod-cloud.net>

aunque sin parar de aumentar la cantidad de paquetes de datos que apuntan a *DBpedia* y al resto de núcleos de flujo; 4) el último diagrama muestra ya los datos clasificados por familias, siendo la más numerosa Publicaciones y la menos, los Contenidos generados por los usuarios.

En estos momentos no contamos con estudios sobre los métodos que pueden utilizarse para analizar este fenómeno. Los diagramas incluidos en este trabajo se han generado sobre registros del *CKAN*, si bien estamos convencidos de que se abrirán líneas de investigación en este sentido con el objetivo de refinar los métodos de análisis.

Sin entrar todavía en el mundo de las bibliotecas, observamos que existen ya grandes depósitos de datos estructurados en *RDF* y listos para ser utilizados por cualquier servicio. Para conocer qué recursos de información han sido publicados siguiendo los cuatro principios del *Linked data* se debe acceder al registro *Datasets in the next LOD cloud*, mante-



Figura 4. Nube de etiquetas de los 1.672 paquetes de datos registrados en *CKAN* (febrero de 2011), <http://ckan.net>

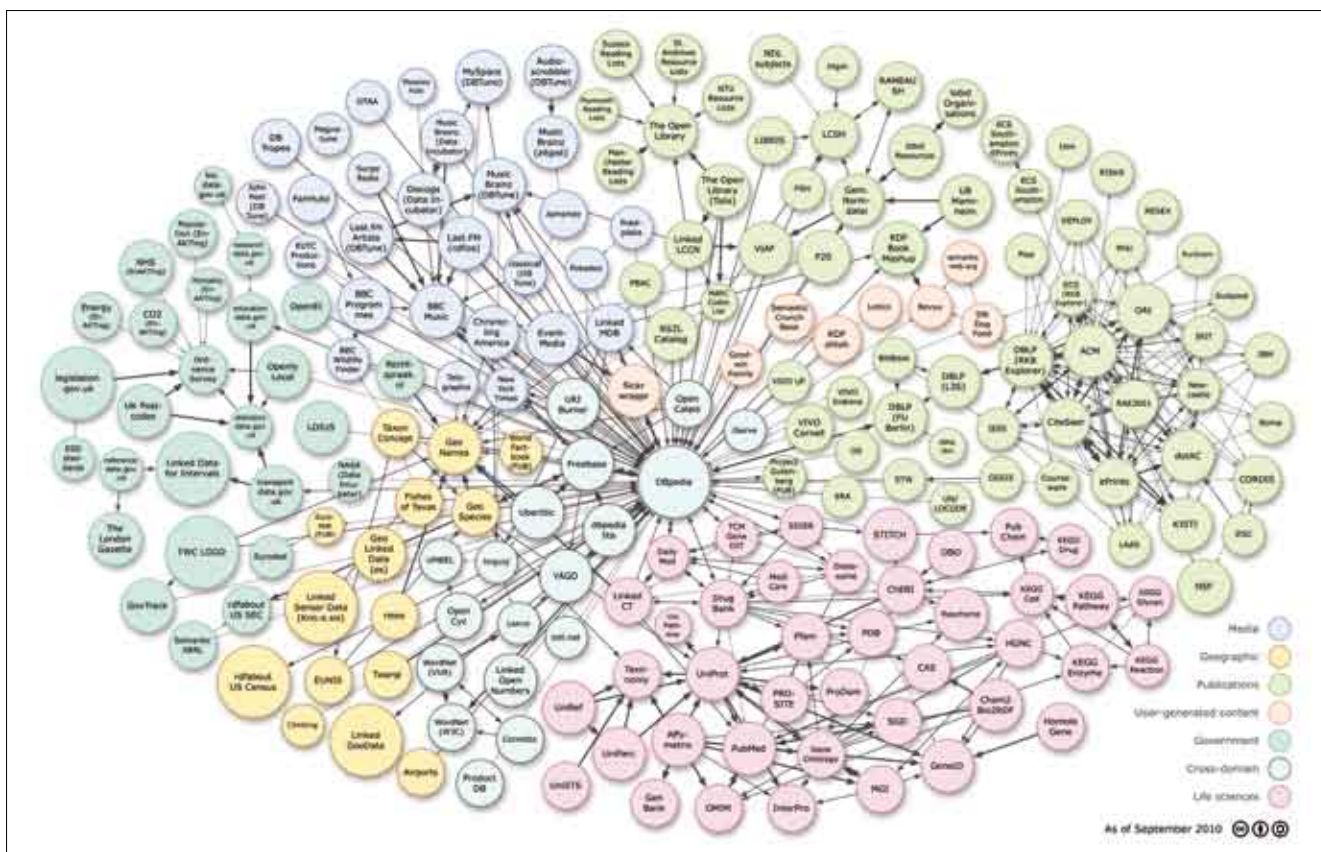
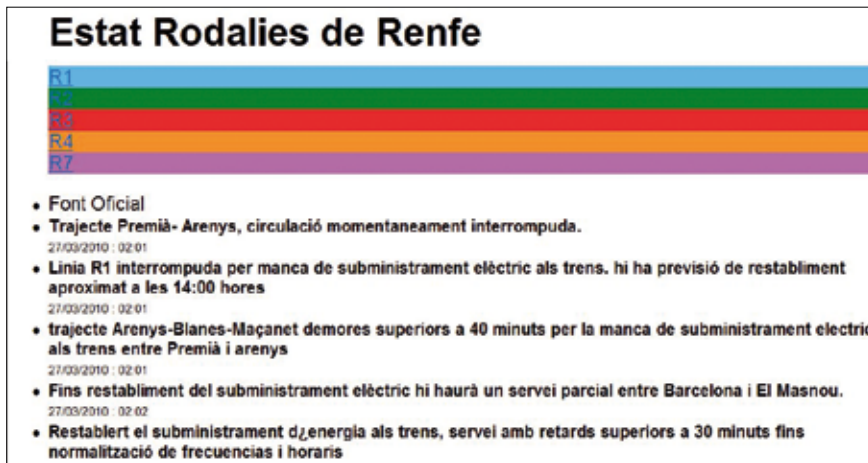


Figura 3. Diagrama de la *Linking data cloud* en 2010, http://richard.cyganiak.de/2007/10/lod/lod-datasets_2010-09-22_colored.html



Figuras 5 y 6. Aplicación realizada mediante tweets (<http://rodalia.info/es>) (arriba) que Manel García reutiliza junto con otras fuentes para crear una web para iPhone en <http://r.odali.es> (abajo)

nido desde la *Freie Universität Berlin*, pues el registro CKAN hace referencia a *open data* en general. A fecha de hoy *Datasets...* contiene 203 elementos, con más de 25.000.000.000 triples RDF totales y mención a sus *outlinks* (enlaces salientes) y sus *inlinks* (enlaces entrantes). Entre ellos destacan: *DBpedia* (se sitúa en el 7º puesto en *datasets*, en el 5º por *outlinks* y en el 11º por sus *inlinks*); la migración de la *BBC* (12º en *inlinks*); así también incluye el proyecto *Genoma*, el conocido proyecto *Gutenberg*, la oficina *Eurostat*, el censo norteamericano (4º en *datasets*) o *PubMed* (8º en *datasets*).

Todavía hay pocos grupos españoles trabajando en esto, y por lo tanto pocos contenidos disponibles en la Web. Pero están ya convertidos a RDF y son referente ciertas bases de datos de la *Universidad Politécnica de Madrid* y algún otro ejemplo, como <http://rodalia.info/es> que ofrece tweets clasificados sobre incidencias en los trenes *Renfe* de cercanías catalanes, que se reutilizan en <http://r.odali.es>. También hay que citar los trabajos de la *Asociación Española de Linked Data (Aelid)*, y del *Ontology Engineering Group (OEG)* de la *UPM* y su proyecto *GeoLinked data*.

Iniciativas relacionadas con bibliotecas

A escala internacional, el W3C identificó la comunidad bibliotecaria como uno de sus focos de interés y a través de sus *Incubator groups* –grupos dinámicos que se utilizan para detectar nuevos nichos de actuación–, creó en mayo de 2010 uno orientado a la publicación de datos estructurados en *Linked data* procedentes de bibliotecas. El *Library*

Linked Data Incubator Group (LLD-XG) tiene como misión analizar la situación de los modelos y esquemas de metadatos y los estándares y protocolos de interoperabilidad que se deberían usar para la publicación y uso de LOD para datos de bibliotecas. A su vez anima a la comunidad bibliotecaria a publicar sus registros bibliográficos y control de autoridades como *Linked data*.

El *LLD-XG* considera las bibliotecas como un proveedor de contenidos con gran potencial. El origen de esta fortaleza se encuentra obviamente en la información bibliográfica de tipo referencial, que se genera mediante normativas consolidadas que aseguran su calidad, capacidad de compartir y sostenibilidad. Además trabajan con vocabularios controlados y con ficheros normalizados de autoridades –nombres de persona, entidades, etc.–, lo que resulta de gran utilidad para enriquecer otros datos y para favorecer el correcto desarrollo de la web semántica.

En septiembre 2010 el *LLD-XG* solicitó estudios de casos, entendidos como iniciativas de divulgación y difusión dirigidas a la comunidad bibliotecaria y los sectores afines. Durante la primera

reunión en noviembre 2010 los casos recibidos se agruparon en seis temas:

- Datos bibliográficos
- Listas de autoridades
- Vocabularios
- Archivos y datos heterogéneos
- Citas
- Objetos digitales
- Colecciones
- Usos sociales

La participación española fue escasa. Además del caso presentado por *Ciepi (IraLIS-International Registry for Authors)*,

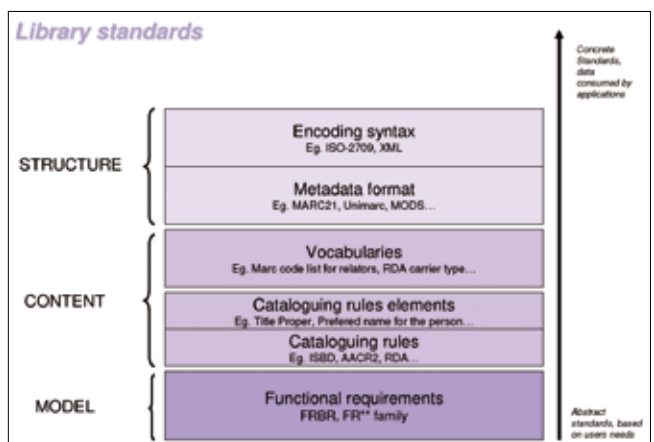


Figura 7. Normativas en bibliotecas identificadas por el LLD-XG <http://www.w3.org/2005/Incubator/llid/wiki/File:LayeredModelV3.pdf>

estuvo la *Biblioteca Virtual Ignacio Larramendi-Colección de polígrafos* y el grupo de vocabularios de la *Fundación Botín*, con la *Ontología cultural de Cantabria*, asociadas al trabajo de Agenjo-Bullón y Hernández-Carrascal (2010), de *Larramendi-Digibis*.

Como muestra de la influencia que va a tener el informe final de *LLD-XG* señalamos que otros casos de estudio los presentaron la *Bibliothèque nationale de France (BnF)*, *Virtual International Authority File (VIAF)*, la *Deutsche Nationalbibliothek (DNB)*, el tesauro *Agrovoc*, la base de datos *Agris*, y *Europeana*, que ha comenzado a migrar los catálogos multimedia de museos y bibliotecas.

En mayo 2011 el *LLD-XG* publicará un informe con los resultados del estudio de la situación que deberá de servir como punto de referencia para que la comunidad de bibliotecarios comience a participar en la web semántica no sólo como usuaria sino como productora de datos.

Estrechamente relacionada con los objetivos de *LLD-XG* pero ampliada a museos y archivos, se ha convocado una cumbre en San Francisco para el 2-3 de junio de 2011 con la intención de reunir representantes clave de sectores del mundo de las bibliotecas, archivos y museos y debatir las posibilidades en estas comunidades, al que asistirá *Ciepi*. Los temas a tratar, de indubio interés, son:

- herramientas y técnicas para publicar y trabajar con *Linked open data*;
- políticas de licencias y copyright;
- publicación de metadatos de bibliotecas, archivos y museos;
- publicación de la terminología y promoción de los casos de uso presentados en octubre;
- promoción de LOD en las instituciones.

Como conjuntos de metadatos ya implementados en la comunidad bibliotecaria destacamos tres proyectos: las autoridades y vocabularios de la *Library of Congress* (EUA); el servicio LOD de la *Deutsche Nationalbibliothek (DNB)*; y *Libris*, el trabajo realizado en el catálogo colectivo de Suecia.

- *Library of Congress. Authorities and vocabularies*
Como parte de sus investigaciones para agregar valor e innovar en nuevos métodos de distribución de la información, la *Library of Congress* ha expuesto sus vocabularios y listas de autoridades en la Web. La más conocida es su lista de materias *LCSH*, pero también ofrece otros recursos sobre lenguajes (ISO 639-1, 639-2 y 639-5 y la Lista MARC), sobre áreas geográficas (*MARC List for geographic areas*) o el tesoro para materiales gráficos (*Thesaurus of graphic materials*).

La asignación de URIs a los valores de sus vocabularios permite aumentar su utilización para otros fines, con una política clara sobre su reutilización, tanto de los valores individuales como de los vocabularios enteros. Lo que hace es proporcionar una interfaz web para la búsqueda humana, pero esta conversión permite además que cuando las máquinas solicitan la URI de un descriptor se carguen datos referentes a otros modelos de descripción como SKOS o JSON, enlaces a otros vocabularios..., como puede verse en el ejemplo de las figuras 8 y 9.

Figuras 8 y 9. Presentaciones textual y gráfica para el término "Library administrators", <http://id.loc.gov/authorities/sh85076660>

En este sentido, no queremos dejar de mencionar nuestro propio trabajo en la conversión a SKOS de *JITA-Classification schema of library and information science*, utilizada en el repositorio *E-prints in Library and Information Science (E-LIS)*. Fue convertido a SKOS recientemente a instancias de **Christophe Dupriez** y publicado en el *Voc-Bench*, herramienta de la *FAO* para la gestión de vocabularios controlados. En breve esperamos contar con varias traducciones por parte del equipo internacional de editores de *E-LIS* y registrarlos como conjunto de datos en el registro *CKAN*.

• *Linked data service* de la *Biblioteca Nacional de Alemania*

Por su parte, el *Linked data service* de la *Deutsche Nationalbibliothek* tiene la intención de exponer todos sus datos bibliográficos y de autoridades, sin perder de vista a medio plazo la necesidad de clarificar en qué términos se ofrece el uso de sus datos a servicios comerciales web. Este enfoque, liderado por **Hanne-mann**, es novedoso en la comunidad bibliotecaria y está basado en las lecciones aprendidas y reflexiones que se suscitaron durante la puesta en marcha del servicio.

Desde esta institución se ha acometido un proyecto realista, limitado a 3,5 millones de datos sobre autoridades (vocabularios-personas-entidades) con el objeto de sentar las bases para posteriores retos. Conecta sus datos con fuentes externas como la *Wikipedia* alemana y *DBpedia*, *VIAF*, *LCSH* o *Rameau* (de la *Bibliothèque nationale de France*). Y utiliza como esquema de datos principal el propuesto en RDA (*resource description and access*).

Esta experiencia ha sacado a la luz otros aspectos que hasta ahora estaban oscurecidos por la concentración de los esfuerzos en las cuestiones técnicas, que abordaremos sucintamente en el siguiente apartado.

• *Libris*, catálogo colectivo sueco

El proyecto sueco constituye una aproximación pragmática a la producción de su catálogo colectivo en formato LOD. En 2009 **Malmsten** presentó unos resultados con el objeto de compartir su experiencia y hacer accesibles los datos, aunque sin la intención de que los metadatos fueran perfectos. Según el registro LOD mencionado anteriormente, con ello han obtenido 5 millones de tripletes con 3 millones de enlaces externos.

Entre sus experiencias y lecciones aprendidas están: desarrollar un API ajeno al sistema bibliotecario –que era comercial–, capaz de extraer de éste los datos necesarios; exportación a MARC XML de los datos, y exportación con URIs de las relaciones entre los datos; construcción de identificadores; transformación del MARC21 a RDF... Abogan por empezar a construir conocimiento al tiempo que se experimenta y se aprende.

Retos para la comunidad bibliotecaria

A tenor de todo lo expuesto no cabe duda que las bibliotecas pueden y deben participar con sus contenidos en la web semántica. Y aquí viene el problema, ¿cómo hacerlo con recursos y conocimientos técnicos limitados? La conversión de recursos a LOD no es sólo un reto técnico, existen otros aspectos (**Hannemann**, 2010) como la falta de experiencia y recomendaciones sobre:

- las propias utilidades tecnológicas para la conversión;
- ontologías y esquemas de datos a utilizar por entidades que ya intercambian datos masivamente con otros formatos;

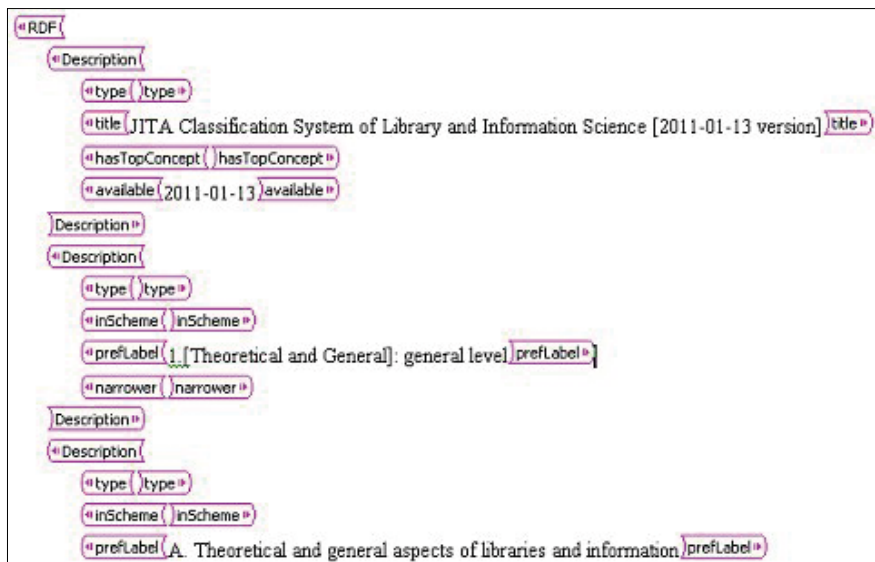


Figura 10. Ejemplo de RDF/XML para JITA

- acuerdos legales y licencias para permitir la utilización de estos datos...; y en general,
- falta de documentación detallada sobre los procedimientos a seguir.

Existe un proyecto que quiere facilitar la publicación de *Linked data* por parte de las bibliotecas y especialmente hacer hincapié en la toma de decisiones en lo que respecta la codificación de los datos. Aunque está orientado al sector de la agricultura, *Virtual open access in agriculture and aquaculture repository (VOA3R)*, puede ser de utilidad para cualquier biblioteca. Se trata de un proyecto europeo que tiene como objetivo el desarrollo de una plataforma para el intercambio de información. Tiene una duración de tres años, 2010-2013, y participan 14 instituciones de 10 países europeos. En el marco de *VOA3R*, y concretamente con el ánimo de facilitar este intercambio, nació la idea de asistir a las bibliotecas para decidir qué propiedades usar en la codificación de los datos bibliográficos y producir *Linked data*. Responde a cuestiones como ¿qué propiedad debo usar para título, o para localización del recurso?, o ¿cómo debo codificar el campo autor?

Para ello se han establecido las *Recomendaciones LOD*, elaborada por **Marcia Lei-Zeng**, *Kent State University*, EUA, e **Imma Subirats-Coll**, *FAO*, que ayuda a los proveedores de datos en la selección de las estrategias más apropiadas para codificar de acuerdo con sus necesidades. Usa diagramas para facilitar la toma de decisiones y escoger las propiedades más usadas en la descripción de los recursos. Las *Recomendaciones LOD* se presentan tanto a nivel gráfico como en texto. Como los proveedores de datos cuentan con diversas estructuras de datos bibliográficos y, por tanto, no hay una solución válida para todos, se ofrecen sugerencias múltiples para la codificación de los datos. Las *Recomendaciones LOD* serán publicadas en mayo 2011, y posteriormente revisadas acorde con las sugerencias recibidas hasta finales de 2011. Aunque nacieron dentro del marco de un proyecto para la creación de un motor de búsqueda como *VOA3R*, el objetivo es que sean aplicables para la comunidad bibliotecaria en general.

Conclusiones

Está claro que el movimiento *Open data* revolucionará en pocos años el mundo del acceso a los datos. En su vertiente vinculada a la administración pública se aprecia ya un movimiento imparable liderado por los propios gobiernos en aras de conseguir su imbricación con el mundo Web. Esta filosofía, que impregna también el segundo aspecto tratado en este trabajo, *Linked open data*, obtendrá sus mejores resultados si la información se suministra además descrita de manera estándar para fomentar la reutilización. Y consideramos que es en este aspecto en el que nuestra comunidad profesional debe estar presente desde un principio, abriendo sus almacenes de datos a otros agentes ajenos a su entorno profesional.

Es indudable la importancia del papel que pueden desempeñar las bibliotecas y otras instituciones que conservan el patrimonio en este escenario de la web semántica, por el uso de programas y normas, por la calidad del trabajo que realizan sus profesionales, y por la estabilidad en los datos de sus catálogos. Para las bibliotecas, a su vez, se percibe la oportunidad de hacer mucho más visible su trabajo y su información, pero también de asumir el papel protagonista que les corresponde. Las instituciones de la memoria ofrecen un mapa congruente de la cultura de los países, algo que actualmente aparece desagregado en la Web.

Compartir datos de forma masiva y estable además ayuda en su propio trabajo, como en la detección de duplicados, la desambiguación terminológica, el enriquecimiento de los datos o en suministrar su información en formatos más transparentes a los usuarios ajenos a nuestra comunidad profesional.

Es el momento oportuno y detectamos una motivación en el sector: es nuestra hora, y debemos participar.

Referencias

Aelid

<http://www.aelid.es>

Agenjo-Bullón, Xavier; Hernández-Carrascal, Francisca. "La Biblioteca Virtual Larramendi: fuente de información bibliográfica para el pensamiento iberoamericano en la Web 3.0". En: *Jornadas virtuales iberoamericanas de bibliotecología*, 2010. http://www.jornadasbibliotecas.bibar.org/index.php?id=9&cid=95&fid=20&task=download&option=com_flexicontent V Congreso nacional de bibliotecas públicas, Gijón (España), 3-5 nov. 2010. <http://eprints.rclis.org/handle/10760/14719>

Agenjo-Bullón, Xavier; Hernández-Carrascal, Francisca. "La construcción de esquemas semánticos para bibliotecas virtuales". En: *I Jornada profesional 'Lenguajes y gestión de información'*, organizada por Sedyc y el Instituto Cervantes, 17 junio 2010. <http://goo.gl/3gMo9>

Alonso, José-Manuel. "Towards a linked government data cloud" (Hacia una nube de datos públicos enlazados). En: *Seminario Visualizar'09: Datos públicos, datos en público*, 12-13 nov. 2009. http://medialab-prado.es/article/seminario_visualizar09_datos_publicos_datos_en_publico

Belanche-Alonso, Xavier. "data404: producir, liberar, re-mezclar y representar la información pública". En: *Seminario Visualizar'09: Datos públicos, datos en público*, 12-13 nov. 2009.

http://medialab-prado.es/article/seminario_visualizar09_datos_publicos_datos_en_publico

Berners-Lee, Tim. *Linked data*, 2006.

<http://www.w3.org/DesignIssues/LinkedData.html>

Biblioteca Virtual Ignacio Larramendi. Colección de polígrafos.

<http://www.larramendi.es/i18n/estaticos/contenido.cmd?pagina=estaticos/bibliotecall>

Bizer, Christian; Cyganiak, Richard; Heath, Tom. *How to publish linked data on the web*, 2007.

<http://www4.wiwiss.fu-berlin.de/bizer/pub/LinkedDataTutorial>

Bizer, Christian; Heath, Tom; Berners-Lee, Tim (en prensa). "Linked data - The story so far". *Intl journal on semantic web and information systems (Ijswis)*.

<http://linkeddata.org/docs/ijswis-special-issue>

<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

Callejero de Gijón.

<http://www.gijon.es/callejero/#0>

Casos de uso. *Library Linked Data Incubator Group*.

<http://www.w3.org/2005/Incubator/ld/wiki/UseCases>

CKAN - the Data Hub.

<http://ckan.net>

Cyganiak, Richard; Jentzsch, Anja. *Linking open data cloud diagram*.

<http://lod-cloud.net>

Datasets in the next LOD Cloud.

<http://www4.wiwiss.fu-berlin.de/lodcloud>

Datos abiertos de la *Generalitat de Catalunya*.

<http://dadesobertes.gencat.cat/es/>

Datos abiertos de Zaragoza.

<http://www.zaragoza.es/ciudad/risp/>

DBpedia.

<http://dbpedia.org/About>

Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 nov. 2003, relativa a la reutilización de la información del sector público. *Diario oficial*, n. L 345, de 31/12/2003, p. 0090-0096.

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32003L0098:ES:HTML>

Domínguez, Eva. "Open data, ¿se abre la caja de Pandora?".

El cuarto bit, 19 nov. 2010. Dónde (en) Zaragoza DND zgz.

<http://www.dndzgz.com/web/index.html>

Dunsire, Gordon; Willer, Mirna. "Initiatives to make standard library metadata models and structures available to the semantic web". *76th IFLA general conf and assembly*, 10-15 August 2010, Gothenburg, Sweden.

<http://www.ifla.org/files/hq/papers/ifla76/149-dunsire-en.pdf>

eGovernment

<http://www.w3.org/egov>

FOAF Friend of a friend.

<http://www.foaf-project.org>

Fundación CTIC.

<http://datos.fundacionctic.org>

GeoLinked Data.

<http://geo.linkeddata.es/web/guest/home>

Guidelines for collecting metadata on linked datasets in CKAN.

<http://esw.w3.org/TaskForces/CommunityProjects/LinkingOpenData/DataSets/CKANmetainformation>

Hannemann, Jan; Kett, Jürgen. "Linked data for libraries". En: *76th IFLA general conf and assembly*, 10-15 August 2010, Gothenburg, Sweden.

<http://www.ifla.org/files/hq/papers/ifla76/149-hannemann-en.pdf>

Incidencias de tráfico en Gipuzkoa.

<http://opendata.blog.euskadi.net/es/demo/trafico-gertae-rak-gipuzkoan/>

Iralis

<http://www.iralis.org>

Iralis. Use case International registry for authors.

http://www.w3.org/2005/Incubator/Ild/wiki/Use_Case_International_Registry_for_Authors

Library Linked Data Incubator Group.

<http://www.w3.org/2005/Incubator/Ild>

Library standards and linked data. Library Linked Data Incubator Group.

http://www.w3.org/2005/Incubator/Ild/wiki/Library_standards_and_linked_data

Linked data - Connect distributed data across the Web.

<http://linkeddata.org>

Linked Data Service.

http://www.d-nb.de/eng/hilfe/service/linked_data_service.htm

Malmsten, Martin. "Exposing library data as linked data", 2009. En: *IFLA 2009 satellite meetings in Florence*.

<http://www.ifla2009satelliteflorence.it/meeting3/program/assets/MartinMalmsten.pdf>

MARC List for geographic areas.

<http://id.loc.gov/vocabulary/geographicAreas>

Mazzo Iturriaga, Rodrigo. Open data: qué es y ejemplos en el mundo, 2010.

<http://www.bcn.cl/de-que-se-habla/open-data-link-data>

Méndez, Eva. "RDF: un modelo de metadatos flexible para las bibliotecas digitales del próximo milenio". En: *7es Jornades catalanes de documentació*, 1999.

<http://www.cobdc.org/jornades/7JCD/1.pdf>

Ontology Engineering Group.

<http://www.oeg-upm.net>

Open Data Euskadi.

<http://opendata.euskadi.net/w79-home/es>

Open Data Foundation.

<http://www.opendatafoundation.org>

Open Government Licence.

<http://www.nationalarchives.gov.uk/doc/open-government-licence>

Open Knowledge Foundation.

<http://okfn.org>

Opening up Government.

<http://data.gov.uk>

Planificador de visitas.

http://www.zaragoza.es/turruta/Turruta/index_Ruta

Portal de datos abiertos de Gijón.

<http://datos.gijon.es>

Predicción Cataluña.

http://dadesobertes.gencat.cat/es/dades-obertes/dataset_000070.html

Proyecto Aporta.

<http://aporta.es/web/guest/index>

Public Sector Information (PSI), Data catalogues (by governments) (Direct access to data).

http://www.epsiplatform.eu/psi_data_catalogues/category_1_public_sector_information_psi_data_catalogues_by_governments_direct_access_to_data Updated: 21 November 2010

Thesaurus of graphic materials.

<http://id.loc.gov/vocabulary/graphicMaterials>

VIAF The virtual international authority file.

<http://viaf.org>

VOA3R.

<http://voa3r.eu>

W3C (2010). Semantic web.

<http://www.w3.org/standards/semanticweb>

Wikileaks

<http://wikileaks.info>