



LINKED DATA Y LINKED OPEN DATA: SU IMPLANTACIÓN EN UNA BIBLIOTECA DIGITAL. EL CASO DE EUROPEANA



Ana Ríos-Hilario, Diego Martín-Campo y Tránsito Ferreras-Fernández



Ana Ríos-Hilario, licenciada y doctora en documentación, es profesora del *Depto. de Biblioteconomía* de la *Univ. de Salamanca*, donde imparte, entre otras, la asignatura de descripción de documentos digitales en el *Máster en sistemas de información digital*. Su labor docente e investigadora se centra en la descripción y acceso documental de los recursos digitales, tema sobre el que ha realizado publicaciones y ponencias. Dirige trabajos de investigación sobre modelos de metadatos y proyectos de datos enlazados. Participa en el *Grupo estándares: sistemas de representación y organización del conocimiento* del proyecto *I+D+i Infoscopos*.

*Departamento de Biblioteconomía y Documentación
Universidad de Salamanca
Francisco de Vitoria 6-16. 37008 Salamanca, España
anarihi@usal.es*

Diego Martín-Campo, diplomado en biblioteconomía y licenciado en documentación por la *Universidad de Salamanca*. Cuenta con estudios de posgrado en tratamiento de información digital (*Máster en sistemas de información digital*), y realiza el doctorado en información y documentación. Trabaja como documentalista en la descripción y difusión de documentos digitales en el *Centro Internacional de Tecnologías Avanzadas* de la *Fundación Germán Sánchez Ruipérez*, y como técnico para el *Repositorio Documental Gredos* de la *Universidad de Salamanca*.

*Departamento de Biblioteconomía y Documentación
Universidad de Salamanca
Francisco de Vitoria, 6-16. 37008 Salamanca, España
diegomartin@usal.es*

Tránsito Ferreras-Fernández es licenciada en filología francesa y máster en sistemas de información digital por la *Universidad de Salamanca (Usal)*. Pertenece al cuerpo de ayudantes de archivos, bibliotecas y museos de la *Usal*. Es coordinadora técnica del *Repositorio Institucional Gredos*. Investiga el impacto del acceso abierto, el desarrollo e implementación de esquemas de metadatos en repositorios *DSpace* y a la preservación digital de repositorios. Participa en grupos de trabajo de proyectos nacionales como *Recolecta*, e internacionales como *Europeana*.

*Universidad de Salamanca
Repositorio Institucional Gredos
Libreros, 68. 37008 Salamanca, España
transiff@usal.es*

Resumen

Se realiza un análisis de los conceptos *linked data* (LD) y *linked open data* (LOD) y su aplicación a un caso real: *Europeana*. Para LD se parte del informe final publicado recientemente por el *Library Linked Data Incubator Group (LLD-XG)*, describiendo seguidamente el caso de *Europeana* y detallando las tareas realizadas y por desarrollar. Posteriormente, se examina la aplicación de LOD a través de la implantación del *Europeana data model* (EDM) haciendo referencia a cada uno de sus componentes. Se concluye con una valoración de la aplicación de estas técnicas en el ámbito de las instituciones culturales.

Artículo recibido el 09-02-12
Aceptación definitiva: 16-05-12

Palabras clave

Linked data (LD), Linked open data (LOD), Library linked data incubator Group (LLD-XG), *Europeana data model* (EDM).

Title: *Linked data and linked open data and its implementation in a digital library: the case of Europeana*

Abstract

An analysis of the implementation of linked data (LD) and linked open data (LOD) techniques in *Europeana* is presented. The final report recently published by the *Library Linked Data Incubator Group* (XG-LLD) is examined, and its application in *Europeana* is analyzed. Details of the work already done and the activities to be developed are provided. Then, the article examines the LOD implementation following the *Europeana data model* (EDM) for each of its components. Finally an assessment on the use of these techniques in the field of cultural institutions is done.

Keywords

Linked data (LD), Linked open data (LOD), *Library Linked Data Incubator Group* (LLD-XG), *Europeana data model* (EDM)

Ríos-Hilario, Ana; Martín-Campo, Diego; Ferreras-Fernández, Tránsito. "Linked data y linked open data: su implantación en una biblioteca digital. El caso de *Europeana*". *El profesional de la información*, 2012, mayo-junio, v. 21, n. 3, pp. 292-297.

<http://dx.doi.org/10.3145/epi.2012.may.10>

1. Introducción: sobre datos abiertos, enlazados y datos abiertos enlazados

Los movimientos *open* (*open access*, *open source*, *open data*) están siendo objeto de atención por la comunidad bibliotecaria en los últimos años, pues aportan numerosos beneficios ante el gran crecimiento de información, ya que aumentan las posibilidades de descubrir, usar y reutilizar los recursos en las instituciones culturales.

Open access se define como el libre acceso a la información y conocimiento a través de internet, sin barreras económicas y sin restricciones derivadas de los derechos de copyright (Ferreras-Fernández, 2011). Las instituciones culturales se benefician del software *open source*, reduciendo los costes iniciales y de mantenimiento al eliminar —en principio— la figura de los proveedores tecnológicos. El software OS facilita a los profesionales un mayor control sobre sus sistemas pudiendo obtener un mejor conocimiento sobre los intereses y expectativas de los usuarios, pues pueden adaptar los programas libremente a sus necesidades. Finalmente, *open data* es un movimiento que promueve la liberación de datos, generalmente no textuales y en formatos reutilizables no propietarios como CSV (*comma separated values*), procedentes de organizaciones (Peset; Ferrer-Sapena; Subirats-Coll, 2011). En este contexto, las bibliotecas juegan un importante papel dentro de la web semántica, al exponer grandes colecciones de datos que los usuarios pueden reutilizar y enriquecer.

Igualmente podíamos hablar de los movimientos "data", entre los que se encuentran *open data*, *linked data* (LD) y la suma de ambos, *linked open data* (LOD). LD (*datos vinculados o enlazados*) se refiere a los datos publicados de acuerdo con los principios diseñados para facilitar los vínculos entre conjuntos de datos, vocabularios controlados y otros elementos (Berners-Lee, 2011). LD utiliza URIs (*uniform resource identifiers*) como identificadores únicos globales para todo tipo de recursos. Los datos enlazados se expresan utilizando estándares como RDF (*resource description fra-*

mework), que especifica las relaciones entre los recursos, las relaciones que se pueden utilizar para navegar entre ellos, o integrar la información de múltiples fuentes.

Por su parte, el objetivo del LOD es la construcción de una base de conocimiento común mediante la publicación en la Web de bases de datos en RDF y el establecimiento de enlaces RDF entre datos de diferentes fuentes.

En este sentido, nos parecen muy oportunas las palabras de Zeinstra y Keller (2011) cuando manifiestan que "los datos pueden estar abiertos pero no enlazados. Los datos pueden estar enlazados pero no abiertos. La web semántica puede solamente funcionar con datos que estén tanto abiertos como enlazados".

Europeana es un portal de internet que actúa como una interfaz de millones de libros, pinturas, películas, objetos de museo y registros de archivos digitalizados de toda Europa (figura 1). Esta biblioteca es también una plataforma para el intercambio de conocimiento que promueve la colaboración entre los agentes culturales.

Los datos pueden estar abiertos pero no enlazados. Los datos pueden estar enlazados pero no abiertos. La web semántica puede solamente funcionar con datos que estén tanto abiertos como enlazados (Zeinstra y Keller)

El objetivo del presente artículo es estudiar los conceptos de LD y LOD a través de un caso real de biblioteca digital: *Europeana*. A su vez, se pretende analizar el informe aprobado en octubre de 2011 por el *Library Linked Data Incubator Group* (LLD-XG, Grupo Incubadora de Datos Enlazados Bibliotecarios) y el modelo de aplicación LOD a través del *Europeana data model* (EDM). Todo ello realizado desde un

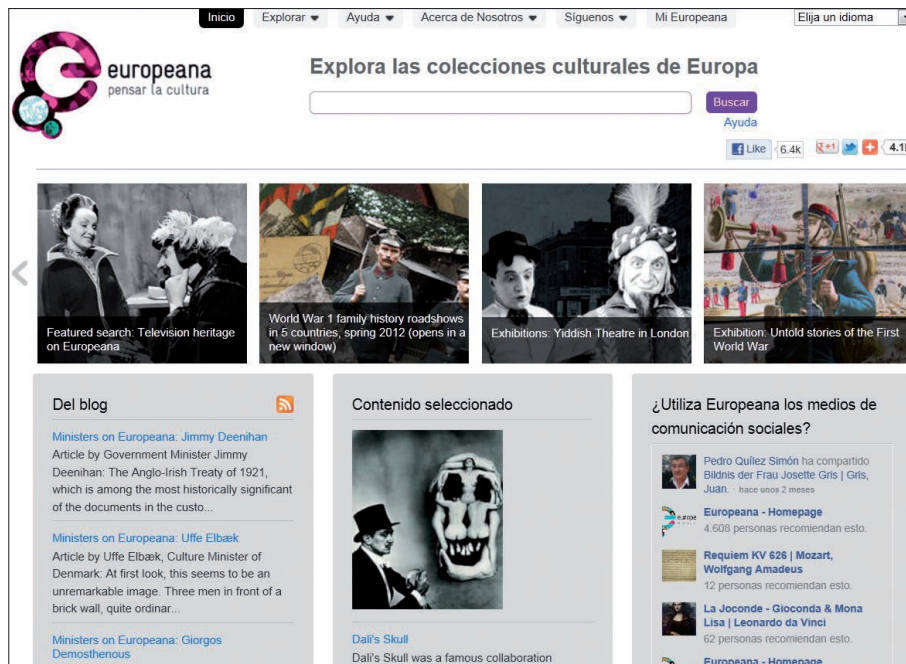


Figura 1. Página de inicio de *Europeana*, <http://www.europeana.eu/portal>

punto de vista eminentemente práctico, centrándonos en la aplicación real de estos conceptos.

2. Library Linked Data Incubator Group (LLD-XG): análisis de uso en *Europeana*

2.1. LLD-XG: Informe final

En mayo de 2010 el *World Wide Web Consortium* (W3C) constituyó el *Library Linked Data Incubator Group*. Su labor consistió en estudiar los componentes básicos ya existentes en este ámbito, tales como los modelos de metadatos y los estándares y protocolos de interoperabilidad, animando a las bibliotecas a publicar sus contenidos y en general a reorientar sus métodos de cara a la interoperabilidad de los datos en la Web. La idea se extendió también a otras comunidades, pues el término *library data* se utiliza en el informe en el sentido más amplio, comprendiendo no sólo los datos bibliotecarios sino también los de archivos y de museos –LAM (*libraries, archives and museums*)–.

El resultado del trabajo del LLD-XG se presentó en octubre de 2011 con el título *Library Linked Data Incubator Group final report*¹, que incluye a su vez otros dos documentos que lo complementan pero que se publican de modo independiente: *Use cases* (“*Usecase*”, casos de usos) y *Datasets, value vocabularies, and metadata element sets* (“*VocabDataset*”, Conjuntos de datos, vocabularios controlados y conjuntos de elementos de metadatos). Según Saorín (2011) este documento “es relevante para entender la oportunidad que se abre a las bibliotecas de hacer mejor las cosas, alcanzar más impacto digital y apostar por formas complementarias de utilidad social de una forma indirecta”.

El cuerpo del documento se divide en los siguientes apartados: ámbito del informe; beneficios de la técnica de datos vinculados; situación actual; y recomendaciones. Resulta interesante la lectura del apartado recomendaciones en función de los distintos destinatarios establecidos.

Una de las principales actividades del *Grupo* fue la recopilación de casos de uso que demuestran la implementación exitosa de las tecnologías de web semántica en las diferentes instituciones culturales. El conjunto de casos prácticos recopilados en el *Usecase* se organizaron en las siguientes agrupaciones:

- datos bibliográficos
- datos de autoridad
- vocabularios
- archivos y datos heterogéneos
- citas
- objetos digitales
- colecciones
- usos sociales

En el siguiente apartado estudiaremos esta publicación a través del análisis detallado de *Europeana*.

VocabDataset es un complemento del informe final que pretende identificar “un conjunto de recursos útiles para la creación o consumo de datos enlazados en el campo de las bibliotecas”. Se concibe como un punto de partida para que los profesionales puedan encontrar, comprender y explorar algunos ejemplos de conjuntos de datos, vocabularios controlados y conjuntos de metadatos.

“ Los estándares de la web semántica y los principios de *linked data* pueden utilizarse para hacer que los valiosos activos de información que la biblioteca crea y mantiene sean más visibles y reutilizables en la Web ”

2.2. Caso de uso: *Europeana*

El LLD-XG incluye al portal *Europeana* dentro del grupo “archivos y datos heterogéneos”. La siguiente figura recoge la estructura del estudio (figura 2). Los apartados de los que consta el análisis son idénticos en cada uno de los casos recopilados por el LLD-XG, con lo cual, a través de este caso específico podremos conocer también el modo de proceder del *Grupo* en la indagación de los casos prácticos. Seguidamente pasaremos a detallar los datos más relevantes obtenidos en el examen de *Europeana*.

Actualmente, la labor de *Europeana* se ve dificultada por la gran variedad de formatos de metadatos utilizados. Es posible exigir a los proveedores presentar los metadatos utilizando un modelo unificado, lo que generalmente se traduce en la pérdida de información útil. También es difícil para la recolección, mantener y explotar los enlaces que conectan los objetos a la información contextual, tanto dentro de las colecciones de objetos agregados por el proyecto como entre estos objetos y otros recursos externos (por ejemplo,

Figura 2. Metadata aggregation for accessing digitized objects, Use Case http://www.w3.org/2005/Incubator/1ld/wiki/Use_Case_Europeana

La recopilación de los casos de uso por el LLD-XG demuestra la implementación exitosa de las tecnologías de web semántica en las bibliotecas y otros sectores afines

Europeana no parte de cero, puesto que para beneficiarse de la técnica de LD ya ha realizado las siguientes tareas: experimentación con la búsqueda semántica RDF para las colecciones, alineamiento de los sistemas de organización del conocimiento y utilización de agregadores de datos que exploran los datos enlazados. Además, este portal emplea modelos de referencia y vocabularios controlados que pueden ser útiles y aplicarse dentro del contexto de los datos enlazados.

3. Aplicación de LOD a través de *Europeana data model* (EDM)

Una de las características fundamentales de *Europeana* es que dispone de un modelo de datos específico, *Europeana data model* (EDM) basado en LOD. Con este modelo y las especificaciones funcionales se inicia el 1 de mayo de 2011 la denominada *fase Danubio* (figura 3) en la que actualmente se encuentra el proyecto (Agenjo-Bullón, 2012). Según estas especificaciones, el objetivo de esta fase es “proveer de una sofisticada funcionalidad a la biblioteca digital *Europeana* con más de 10 millones de ítems procedentes del patrimonio cultural y científico de Europa”.

El deseo de *Europeana 2011* que era llegar a ser una institución distribuidora, facilitadora e innovadora (Bloomberg, 2010) no hay duda de que ya se está cumpliendo.

Wikipedia). La reutilización de los objetos y sus datos adjuntos del mismo modo se ve obstaculizada por la rigidez de los medios para acceder a ellos (portal y API –*application programming interface*–).

La utilización de LD y otras técnicas de web semántica pueden aplicarse en *Europeana* para recolectar objetos mejor contextualizados, enriquecer los metadatos ya existentes, y para proporcionar un acceso directo a los mismos lo que permite su reutilización por terceros. Igualmente, su uso posibilita la coexistencia de varios niveles de generalidad en una misma base de conocimiento y una mejora en los procesos de búsqueda.

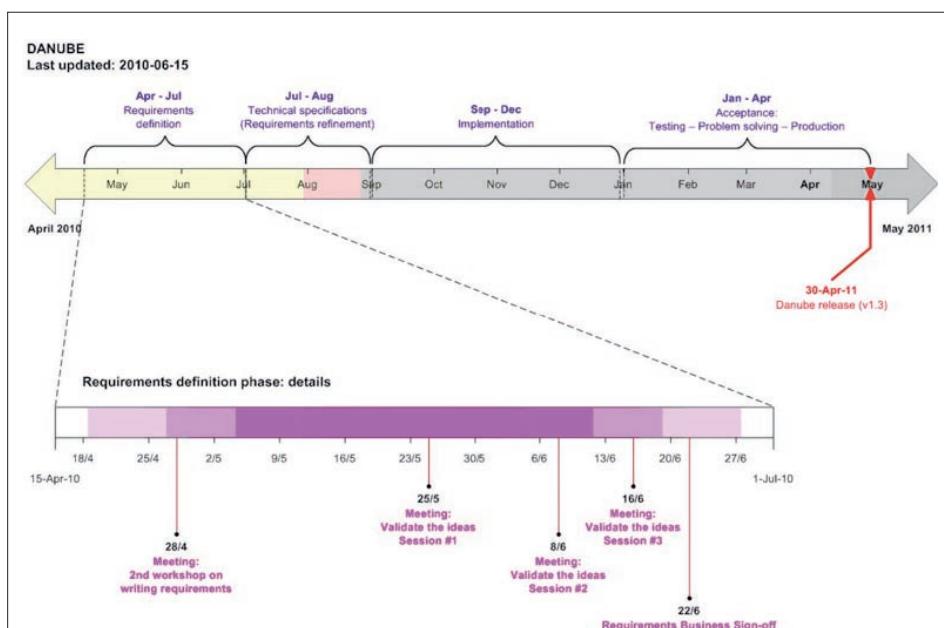


Figura 3. Cronograma de la fase Danubio http://version1.europeana.eu/c/document_library/get_file?uuid=42fdf604-8bd7-4fe4-8d3e-36cce6fcad65&groupId=10602



Figura 4. *Europeana data model*

La figura 4 sintetiza el cambio del modelo *Europeana semantic elements* (ESE) a EDM.

Componentes de *Europeana data model*

Europeana pretende ofrecer una contextualización semántica rica en las representaciones de objetos, para realizar sobre ellos operaciones semánticas complejas que no podrían ser soportadas en la interfaz tradicional de una biblioteca digital. Para ello las representaciones de objetos de esta biblioteca necesitan estar conectadas sistemáticamente a LOD, bien en internet o en recursos de contextualización semántica propios de *Europeana*, como tesauros y vocabularios estructurados migrados al estándar SKOS (*Simple knowledge organization system*). Por tanto, la estrategia técnica de la institución fue concebida para contribuir al paradigma LD y, al mismo tiempo, beneficiarse de él.

La adopción de EDM permite a *Europeana* ser compatible con la web semántica utilizando los estándares del W3C. EDM está constituido por la reutilización de *namespaces* como RDF, RDFS (*resource description framework schema*), OAI-ORE (*Open archives initiative object reuse and exchange*), SKOS y DCterms (*Dublin core metadata initiative terms*).

LD utiliza la arquitectura de la Web como estructura principal y RDF como lenguaje de representación. Dicho estándar permite representar información estructurada acerca de cualquier recurso mediante simples declaraciones triples (sujeto, predicado, objeto), es decir, los recursos se describen por medio de conexiones entre ellos. Por otra parte, EDM permite convertir cualquier objeto de interés en un recurso, identificándolo mediante una URI.

El enfoque de web semántica permite la combinación de varias ontologías en las mismas descripciones, entendidas esas ontologías como conceptualizaciones (Gruber, 1993). EDM reutiliza algunas de las ontologías de referencia ya existentes, como el estándar SKOS. En esta línea, ya se ha publicado la *Lista de encabezamientos de materia para bibliotecas públicas* que

establece un *closeMatch* con los *Library of Congress subject headings* (LCSH) y recientemente con *Rameau (Répertoire d'autorité-matière encyclopédique et alphabétique uni-fié)* y los *Schalgwortnormdatei* (Agenjo-Bullón, 2012).

En cuanto al esquema de descripción *Dublin Core* (DC) proporciona un modelo compacto para describir las características esencia-

les de objetos culturales o científicos (creadores, relaciones con otros recursos, indización por materias, etc.) ajustándose a un amplio rango de necesidades de acuerdo con la web semántica. DC se utilizó también como base del esquema ESE, lo que permite a los proveedores de datos seguir utilizándolo en su versión 3.4., para enviar sus datos a *Europeana*. DC se emplea en muchas aplicaciones permitiendo así que los datos en EDM sean más susceptibles de compartir y reutilizar.

Para el intercambio y reutilización de objetos, se han escogido las especificaciones de OAI-ORE (figura 5) como marco estructural para el modelo de la ontología EDM, ya que, las representaciones de objetos típicas de *Europeana* son entidades compuestas de varias partes, como por ejemplo metadatos de atributos, una vista en miniatura y una página estática en html. OAI-ORE permite un enfoque para la identificación y descripción de estos conjuntos de recursos web.

4. Conclusiones

La aplicación de las técnicas de LD y LOD en las instituciones culturales es ya un hecho, tal y como recogen los casos de uso recopilados por el LLD-XG. Además, el informe de dicho grupo contribuye a sentar las bases en la materia a través de sus recomendaciones.

La implementación de LD en las bibliotecas como *Europeana*, va a permitir un enriquecimiento en todos sus niveles tanto de los datos, modelos de descripción y procesos de búsqueda. Así, y a tenor del trabajo ya realizado, podemos concluir que esta gran biblioteca digital se está convirtiendo

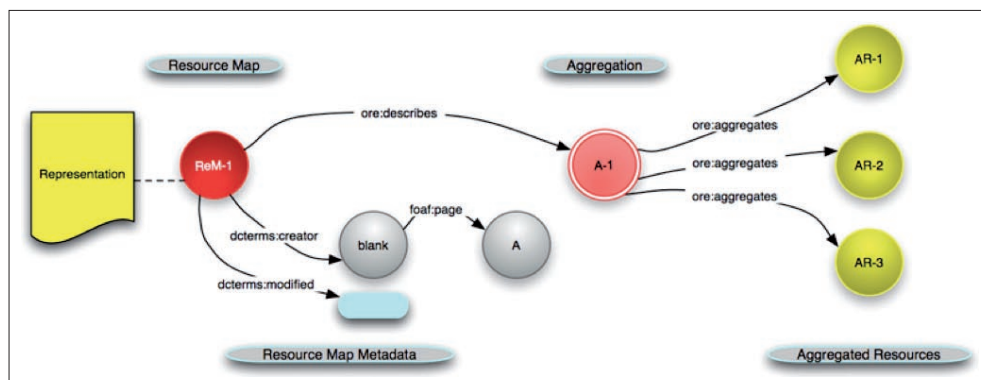


Figura 5. Ejemplo de agregación con OAI-ORE, <http://www.openarchives.org/ore/1.0/primer>

en abanderada de la implementación tecnológica de LD y LOD.

Del estudio realizado extraemos las siguientes apreciaciones sobre la aplicación de dichas técnicas en el entorno de las instituciones culturales. Existen esquemas y modelos de metadatos que evolucionan hacia la Web, normas y protocolos para la construcción de la interoperabilidad más allá del ámbito bibliotecario que cambian de una visión integrada a un entorno de red. Una gran cantidad de datos estructurados ya están disponibles en los sistemas bibliotecarios y podrían publicarse como LD utilizando las técnicas de web semántica. Las instituciones del patrimonio cultural podrían ser un importante proveedor de bases de datos de autoridad (personas, materias...) para la web de datos enlazados.

Sin embargo, también hay que salvar una serie de obstáculos: la adopción real de los principios mismos de datos enlazados en estas comunidades tendrán que coordinarse. Por otra parte, LD a menudo se está construyendo a un nivel de investigación y desarrollo, y trasladar sus resultados a la práctica parece difícil (*Library Linked Data Incubator Group final report*, 2011).

Recientemente **Tomás Saorín** (2011) planteó algunas cuestiones sobre LOD y las oportunidades que representa este nuevo paradigma para las bibliotecas. En este sentido afirmaba “es momento de acordar cuál ha de ser la posición en la que las bibliotecas pueden ser más útiles y eficientes en el mundo *linked open data*”. Nosotros consideramos que es una oportunidad para nuestro sector. Es hora de ponerse manos a la obra.

5. Nota

1. El documento *VocabDataset* fue traducido al castellano primero por Saorín y Pastor:
<http://skos.um.es/Incubator/llid/XGR-llid-vocabdataset>

En marzo de 2012 se publicó la versión española de todos los documentos del informe realizada por Agenjo y Hernández.
http://www.larramendi.es/i18n/cms/elemento.cmd?grupo=estaticos/paginas&elemento=documentacion_tecnica_lam.html

6. Referencias

Agenjo-Bullón, Xavier. “*Europeana* en fase Danubio”. *Blok de BiD*, 25 enero 2012.
<http://www.ub.edu/blokdebid/es/node/157>

Berners-Lee, Tim; Fielding, Roy; Masinter, Larry. *RFC 3986 - Uniform resource identifier (URI): generic syntax*. The Internet Society, 2005.

<http://tools.ietf.org/html/rfc3986>

Berners-Lee, Tim. *Linked data*. World Wide Web Consortium, 18 junio 2009.
<http://www.w3.org/DesignIssues/LinkedData.html>

Bloomberg, Ruth; Dekkers, Makx; Gradmann, Stefan; Lindquist, Mats; Meghini, Carlo; Verleyen, Julie. *Functional specification for the Europeana Danube release*. Europeana Think Culture, 2010.
http://pro.europeana.eu/documents/844813/851994/D3_2_Final.pdf

Europeana Semantic Elements, v.3.4, 2011-03-31.
<http://www.europeana.eu/schemas/ese>

Europeana
<http://www.europeana.eu>

Ferreras-Fernández, Tránsito. “Open access en España: los repositorios institucionales, 2011”. En: *V Jornadas de e-learning en la formación para el empleo en las administraciones públicas*, Valladolid, 15-17 sept. 2011.
<http://eprints.rclis.org/handle/10760/16355#.T7evSZ91Dsw>

Gruber, Thomas R. “Toward principles for the design of ontologies used for knowledge sharing”. *Intl journal human-computer studies*, 1993, n. 43, pp. 907-928.
<http://tomgruber.org/writing/onto-design.pdf>

Library Linked Data Incubator Group final report.
<http://www.w3.org/2005/Incubator/llid/XGR-llid-20111025>

Marcos-Martín, Carlos; Soriano-Maldonado, Salvador-Luis. “Reutilización de la información del sector público y open data en el contexto español y europeo. *Proyecto Aporta*”. *El profesional de la información*, 2011, mayo-junio, v. 20, n. 3, pp. 291-297.
<http://dx.doi.org/10.3145/epi.2011.may.07>

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. 165-173.
<http://www.elprofesionaldelainformacion.com/contenidos/2011/marzo/06.pdf>
<http://dx.doi.org/10.3145/epi.2011.mar.06>

Saorín, Tomás. “Cómo *linked open data* impactará en las bibliotecas a través de la innovación abierta”. *Anuario ThinkEPI*, 2012, v. 6, pp.
<http://www.thinkepi.net/como-linked-open-data-impacta-ra-en-las-bibliotecas-a-traves-de-la-innovacion-abierta>

Zeinstra, Maarten; Keller, Paul. *Open linked data and Europeana*.
http://pro.europeana.eu/c/document_library/get_file?uuid=374c381f-a48b-4cf0-bbde-172cf03672a2&groupId=10602