



VII Encuentros de Centros de Documentación de Arte Contemporáneo

Departamento de Biblioteca y Centro de Documentación del Museo Artium

29 y 30 de octubre de 2014. Museo Artium
Liderazgo de los servicios de información en el siglo XXI

**En el Reino de LOD-LAM... Linked Open Data y
Patrimonio Digital**

Ponentes: Eva M. Méndez Rodríguez y Gema Bueno de la
Fuente

Departamento de Biblioteconomía y Documentación de la
Universidad Carlos III de Madrid

En el Reino de LOD-LAM... Linked Open Data y Patrimonio Digital

1. Introducción

Para hablar del «liderazgo de los servicios de información en el siglo XXI», es necesario que estos estén incluidos de una forma prominente en la Web. Después de 25 años de Web –y, para algunos, 15 de Web Semántica–, la World Wide Web ya no es un lugar «en el que hay que estar», sino que «hay que estar de una determinada manera», en el que «hay que ser». Ser Web es lo mejor que le puede pasar a los servicios de información, entendiendo por tales toda suerte de Bibliotecas, Archivos, Museos y Galerías, que últimamente tienen a bien responder a la denominación afortunada de sus siglas en inglés LAM (Libraries, Archives and Museums) o GLAM, incluyendo también las galerías.

Nunca hemos tenido el patrimonio cultural global tan cerca. En los últimos diez años se ha invertido mucho esfuerzo y dinero, tanto por parte de instituciones públicas, como privadas, en la digitalización de colecciones y contenidos, así como en la creación de millones de registros de metadatos que describen dichas colecciones y objetos digitales del nuevo patrimonio cultural. Sin embargo, aún queda mucho por hacer para la interoperabilidad de dichas colecciones, para las búsquedas simples, y para recuperar información a través de diversas colecciones y servicios de información digital de forma transparente, integrada y eficaz. Las instituciones de la memoria, los/las GLAM, aspiran a convertir sus colecciones digitalizadas más en una experiencia social que en una muestra de contenido. Es crucial que las colecciones sean capaces de pasar de un mero nivel de exposición y acceso a un nivel de reutilización, fuera del museo, de la biblioteca o del archivo.

Hablar de LOD o de Linked Open Data es algo ya habitual en un contexto informativo Web, pero aquí nos gustaría hablar de un nuevo paradigma, de un nuevo espacio, de un nuevo reino. Las/los GLAM llevan muchos años creando metadatos, con la tácita y noble promesa de mejorar la interoperabilidad, esa palabra mágica, que muchas veces no sabemos qué implica entre los distintos servicios de información en la Web. Sin embargo, la calidad de esos metadatos, o el nivel de exposición o disponibilidad de los mismos, no es siempre el que hubiéramos soñado. La implementación de distintas tecnologías, estándares y protocolos ha dispersado los metadatos de nuestras bibliotecas, archivos y museos, en distintas bases de datos, hojas de cálculo, e incluso en documentos textuales no estructurados. Los metadatos de muchas de esas colecciones tienen distintos niveles de detalle y calidad, están en muchos casos formando silos de datos o islas semánticas (semánticas pero al fin y al cabo islas). Linked Data (LD) entraña la nueva promesa de enlazar o crear vínculos entre objetos de diversas colecciones dispares, aunque su implementación real envuelve todavía cierta complejidad, y se precisa entender cómo funcionan los estándares de la web semántica (RDF, SPARQL, etc.) que nos permitirán dotar de esa nueva dimensión «enlazada» a nuestros objetos de información digital.

A pesar de toda esta dificultad, y de la maraña de datos en la que puede convertirse el reino LOD, nos gusta hablar de Linked Open Data como el nuevo reto para los servicios de información y para el patrimonio cultural digitalizado, como un cambio top-down, ya que son muchas las instituciones de la memoria que están abrazando esta nueva tendencia. Así pues, a lo largo de esta publicación analizaremos algunos casos de uso y aplicación real de LOD en archivos, bibliotecas y museos, así como algunos de los principios y fundamentos que permitirán que los datasets de las instituciones de la memoria, recogidos hoy en silos o islas semánticas particulares, pasen a formar parte de la galaxia LOD, y estén disponibles y sean reutilizables como parte fundamental de la Información del Sector Público (PSI).

2. Evolución y madurez de la Web Semántica.

Como decíamos antes, la Web existe desde hace 25 años; el Dublin Core y los metadatos, desde hace 20. La idea de poner semántica a la Web no es nueva tampoco, han pasado más de 15 años desde que comenzó a discutirse el estándar RDF (Resource Description Framework). Quince años son muchos años para seguir hablando de la web semántica como la web del futuro. Lo que antes era una promesa para los servicios de información digital, que además nos llevaría a una «web más bibliotecaria» (Méndez, 2004), es un proceso inminente: el paso de la web de documentos a la web de datos.

Para lograr el desarrollo de la web de datos enlazados, las tecnologías deben permitir que los datos estén disponibles en un formato común (RDF, Resource Description Framework), generados directamente en este formato o mediante la conversión a partir de bases de datos existentes (relacionales, XML, HTML, etc.).

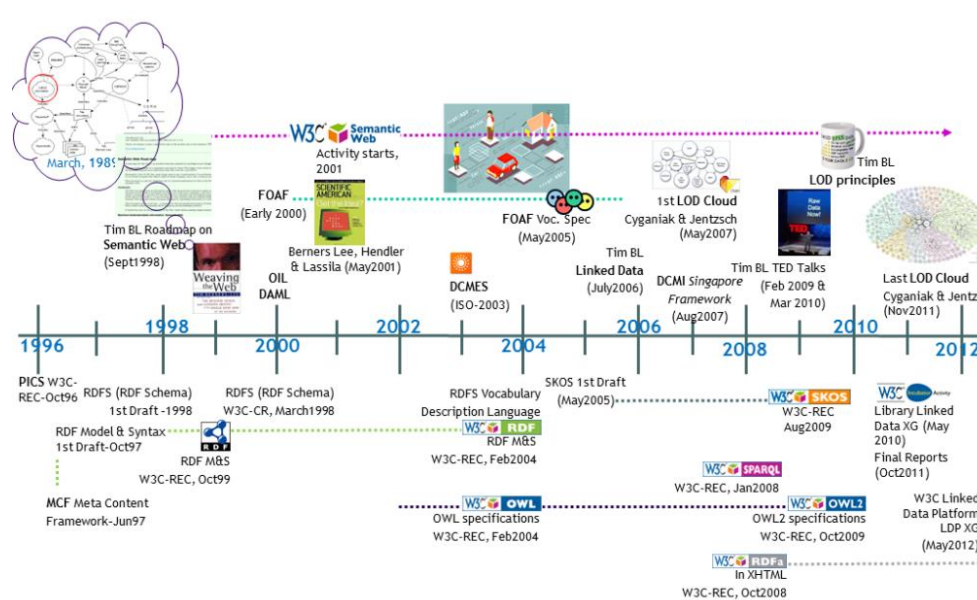


Figura 1. 15 años de web semántica: evolución de los estándares (Mendez, Greenberg, 2012)

3. Fundamentos de Linked Open Data: estándares y otros compañeros de viaje.

3.1. Estándares para la Web Semántica

En el apartado interior hemos dejado patente que, en cierto sentido, la evolución de la Web es la evolución de sus estándares y, sobre todo, que los pilares sobre los que se funda Linked Open Data y la web semántica continúan siendo un conjunto de estándares o especificaciones de dominio público, con vocación de estándar formal. Concebidos, en la mayor parte de los casos, como Recomendaciones del W3C. RDF, GRDDL, POWDER, RDFa, R2RML, RIF, SPARQL, son algunas de las siglas que albergan un conjunto de normas técnicas y tecnologías, que permiten publicar y acceder a los datos que conforman esta web.

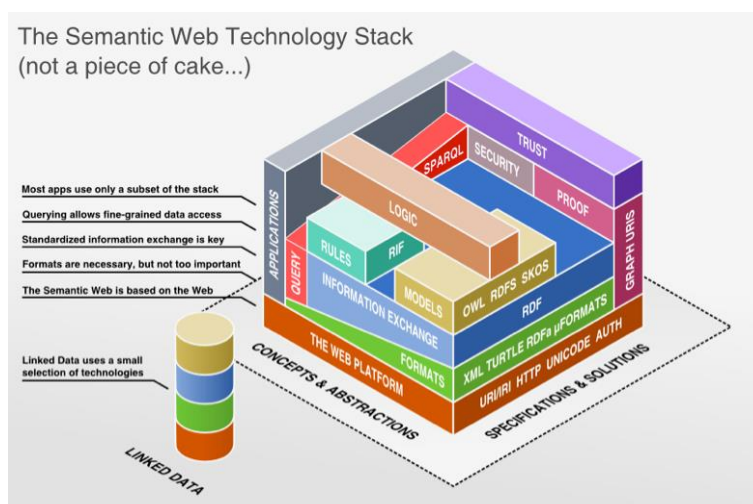


Figura 2. Representación gráfica del conjunto de tecnologías de la Web Semántica (Fuente: http://bnode.org/media/2009/07/08/semantic_web_technology_stack.png)

Estas tecnologías se representan generalmente como una paquete con distintas capas (fig. 2), en la que las especificaciones y estándares se organizan en torno a los conceptos y abstracciones relativos a: la plataforma o arquitectura web, los formatos, el intercambio de información, los modelos de datos, la consulta, y otros aspectos clave para la localización, recuperación y acceso a los datos, mediante la definición de reglas de inferencia, la lógica, la seguridad, la protección y la confianza.

Algunas de las especificaciones más importantes son:

- A nivel de plataforma/arquitectura web: los URIs (Uniform Resource Identifiers) o IRIs (Internationalized Resource Identifiers), estándares del IETF (Internet Engineering Task Force) para la identificación de recursos. Un aspecto fundamental que cobra aún más importancia en el paradigma de la Web de Datos y LOD es la identificación persistente de los recursos, a lo largo de todo el proceso de creación, conversión, publicación, localización, acceso y consumo de los datos. Estas URIs deben ser, además diferenciables, por lo que deben emplear otro estándar básico de la web, el protocolo

HTTP (HyperText Transfer Protocol). Las URIs HTTP son consideradas nombres de los datos, no direcciones web.

- En cuanto a formatos: lenguajes de marcado como XML (eXtensible Markup

Language) o el marco de descripción de recursos en atributos, RDFa (Resource Description Framework—in—attributes) son formatos que permiten estructurar y embeber datos en los documentos en la web.

- En relación al intercambio de información: el marco de descripción de recursos RDF (Resource Description Framework) permite estandarizar la representación de los datos de forma significativa y comprensible para las máquinas, además de para las personas, haciendo posible el intercambio de conjuntos de datos entre distintas plataformas y sistemas. Estos conjuntos de datos en RDF se almacenan en forma de triples, en los llamados triple store, para permitir su consulta y/o descarga.
- Para la consulta: es necesario un lenguaje de consulta común, y un protocolo que regule las consultas a los almacenes de triples RDF, para lo que se ha desarrollado SPARQL (Simple Protocol and RDF Query Language).
- Y para el modelado de los datos: las especificaciones para la definición de ontologías como OWL (Ontology Web Language) y otros vocabularios o sistemas de organización del conocimiento, como SKOS (Simple Knowledge Organization System), permiten el modelado de los datos mediante la definición de clases y propiedades, así como de las relaciones entre estas.

3.2. El paradigma de los metadatos

La comunidad de las bibliotecas y del patrimonio digital ha puesto un gran esfuerzo en los últimos 20 años en la normalización de esquemas de metadatos y vocabularios semánticos, así como en la creación de inmensas colecciones de metadatos. Pensemos por ejemplo en Europeana, que recoge hoy (2014), casi 36 millones de registros de metadatos que describen otros tantos objetos de información digitalizados de las diferentes instituciones de la memoria europea, y que utilizan esquemas de codificación de vocabularios de distinta naturaleza (tesauros, ontologías, etc.)

Tal y como apuntan Seth van Holland y Ruben Verborg en su reciente libro (2014), la evolución de una información no estructurada a una representación de metadatos altamente estructurados requiere el desarrollo de schemas para hacer los metadatos interoperables. Para convertir recursos no estructurados en campos bien estructurados, tenemos que hacer que la semántica de esos campos (los atributos) esté bien documentada en un esquema. Estructurando y atomizando los campos de metadatos los hacemos más interoperables para las máquinas, pero también los hacemos cada vez más dependientes del esquema. Además, cada comunidad, cada dominio o proyecto que se precie en el entorno digital, desarrolla su esquema de metadatos, su vocabulario o vocabularios o, en su defecto, la adaptación de un esquema preexistente, que es lo que conocemos como perfil de aplicación. De tal forma, que cada vez nos resulta más difícil interpretar los metadatos de otro proyecto o sistema de información digital, si no conocemos su schema. Esto hace que los metadatos sean muchas veces como los cepillos de dientes: todo el mundo piensa que son una buena idea, pero cada uno quiere utilizar el suyo.

Cada vez son más las comunidades (dentro y fuera de los GLAM) que crean y utilizan esquemas (schemas) de metadatos como el Dublin Core, o estándares y perfiles de aplicación de metadatos particulares. Estos colectivos o comunidades crean sus estándares descriptivos, que recogen conjuntos de elementos y propiedades (atributos) para conformar sus vocabularios de metadatos. Asimismo, la comunidad de la web semántica del W3C, desde hace menos de un año, subsumida en la comunidad de datos, como dijimos, sigue desarrollando y utilizando ontologías y vocabularios formalizados. Utilizan OWL (Web Ontology Language) para construir dichos vocabularios y ontologías, y SKOS para crear sistemas de organización del conocimiento (KOS), como parte de un complejo proceso para el necesario enriquecimiento de los datos, en el nuevo entorno informativo. Incluso los desarrolladores web han vuelto a resucitar los metadatos vinculados a las páginas web, incorporando datos estructurados que describen o anotan el contenido de documentos HTML. Para estas descripciones embebidas en el código fuente de un documento html utilizan estándares de codificación como microformatos, microdatos y RDFa, y usan también su vocabulario o sus metadatos compartidos, como Schema.org.

La creación de esquemas de metadatos y otros vocabularios son, en cualquier caso, parte de la gestión de la información digital, y todos estos vocabularios constituyen ya un paradigma habitual.

3.3. El paradigma LOD: Principios y Fundamentos de Linked Open Data

En los últimos 20 años, gracias a las tecnologías o, mejor dicho, gracias a Internet, venimos escuchando muy frecuentemente, quizás demasiado frecuentemente, a nuestro juicio, que estamos «ante un nuevo paradigma» o que tal o cual comportamiento tecnológico o tecnología, implica un cambio de paradigma. Sin embargo, lo particular de un paradigma es su carácter fundamentante, desde un punto de vista epistemológico, que conforma un nuevo marco teórico o conjunto de teorías. Así, los requisitos de los datos abiertos enlazados o Linked Open Data (LOD), tal y como los define Tim Berners-Lee en su esquema incremental de cinco estrellas (Berners-Lee, 2006), sin formularlo como tal (un nuevo paradigma), establece las condiciones necesarias para que los datos respondan a un nuevo modelo. Así, para que podamos hablar de LOD los datos deben estar publicados en la web, esto es, tengan un URI, que estén estructurados de forma legible por máquina, en formatos no propietarios, conforme a estándares del W3C, y enlazados con otros conjuntos de datos. Además, estos requisitos, se explicitan de forma más técnica como los cuatro principios de Linked Data (Heath y Bizer, 2011):

1. Usar URIs como nombres para las cosas.
2. Usar URIs HTTP, de modo que la gente pueda buscar y localizar esos nombres.
3. Cuando alguien busque empleando un URI, proporcionar información útil, utilizando estándares (RDF, SPARQL).
4. Incluir enlaces a otros URIs, para que se puedan descubrir y localizar más cosas.

4. Instituciones GLAM y la reutilización de la información en el sector público

Inicialmente se empezó a hablar de LOD-LAM (<http://lodlam.net>) para referirse a la aplicación de Linked Open Data a las instituciones de la memoria (bibliotecas, archivos

y museos) en 2010, acuñado por Jon Voss. Desde entonces, se considera una red de entusiastas, técnicos, profesionales y básicamente, «freakies», que están interesados en trabajar con LOD relativos a galerías, bibliotecas, archivos y museos. De ahí que también se hable de GLAM al hablar de Galerías. Para nosotras, las/los GLAM han sido siempre en entorno perfecto para Linked Open Data: Linked se refiere a conectar conceptos, etc. utilizando los estándares del W3C; Open se refiere al uso de licencias abiertas para contenidos de las instituciones de la memoria, por ejemplo como CC0 (Creative Commons Public Domain Mark, etc.); y Datos son, básicamente los «raw data» de Tim Berners-Lee, o bien metadatos, datos descriptivos, datos bibliográficos, vocabularios, y otros datos que, en el reino LOD-[G]LAM, podrían ser fácil, aunque necesariamente, abiertos, enlazados y compartidos. Aunque esto es un hecho para nosotras, y una realidad para la comunidad LODLAM que organizó su primer workshop en 2011, tendríamos que esperar hasta el reconocimiento de los/las GLAM como parte de la información del sector público (PSI) susceptible de ser abierta y, aún más, enlazada, según los parámetros del nuevo paradigma.

La reutilización de la información en el sector público (RISP), entendida como «el uso por parte de personas físicas o jurídicas, de información generada por organismos del sector público, con fines comerciales o no» (Proyecto Aporta, 2009) es algo latente y vinculado a la transparencia y a la noción de «lo público». Sin embargo, esta reutilización se ha potenciado o propiciado por el desarrollo tecnológico, no solo en cuanto a las posibilidades de acceso e intercambio de información que ofrecen las TIC, sino, especialmente, por su evolución en las capacidades de análisis, explotación y tratamiento de datos, que permite la creación de servicios y aplicaciones basados en el uso, la agregación o la combinación de datos.

Esta reutilización se promueve desde las administraciones públicas a instancia del propio Parlamento Europeo y del Consejo, que con su Directiva 2003/98/CE pretende catalizar un cambio de cultura a favor de la reutilización de la gran variedad y cantidad de información, fiable y de calidad, que producen y gestionan los organismos públicos. En España, este impulso se ha plasmado en la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, que se desarrolla en el Real Decreto 1495/2011, y que, entre otros resultados, ha supuesto la creación de portales de datos abiertos como datos.gob.es a nivel estatal, y otros a iniciativa de diversos gobiernos de comunidades autónomas, ayuntamientos y otras entidades.

La iniciativa de revisar esta directiva por parte de la Comisión, primero en 2011 (Comisión Europea, 2011), y después en 2013, completa el vacío que implicaba la exclusión de las instituciones de la memoria en la versión inicial, y amplía su ámbito de aplicación a las bibliotecas (incluidas las universitarias), archivos y museos. Se reconoce así que los materiales de archivo, colecciones de libros, y obras de arte constituyen también parte del patrimonio de información y contenidos de los organismos públicos. La complejidad y riqueza del universo de bibliotecas con un compromiso público representa una gran diversidad y tipología de colecciones y de datos potencialmente reutilizables.¹

¹ Cabe destacar, no obstante, que no todos los fondos de las bibliotecas pueden ser abiertos y reutilizados, ya que la legislación no se aplica a aquellos documentos y datos con derechos de propiedad intelectual, aunque estén en la biblioteca. Así se ha insistido en la revisión de la directiva publicada el 26 de junio de 2013 (Parlamento Europeo, Consejo, 2013).

Asimismo, la Comisión Europea publicó en Julio 2014 una Comunicación que insta a los ciudadanos europeos a una Economía de Datos próspera, en la que se definen los datos, según la ISO/IEC 2382-1, como una «representación reinterpretable de información de una manera formalizada, adecuada para la comunicación, la interpretación o el procesamiento» (Comisión Europea, 2014). Algunos ejemplos son: información geoespacial, estadísticas, datos meteorológicos, datos de investigación, etc., y también, nuestros datos bibliográficos y todos los datos de distinta naturaleza que, cada vez más, se procesan, comunican, representan o reutilizan, en los sistemas y servicios de información GLAM. Algunos ejemplos pueden observarse en los casos de uso analizados por el Library Linked Data Incubator Group del W3C (W3C Incubator Group, 2012).

Aunque desde hace ya décadas se viene hablando de un cambio de paradigma para las bibliotecas, de la tradición en papel a la biblioteca digital, del átomo al bit, ahora más que nunca esta transición se está haciendo realidad. Como afirma Stefan Gradmann en su ponencia From Containers, to Content, to Context (Gradmann, 2012), las bibliotecas deben pasar de ser contenedoras a contextualizadoras de la información. Es una oportunidad que no pueden perder si no quieren correr el riesgo de volverse insignificantes. Gradmann apunta las claves de esta transición comenzando con la necesidad de cambiar la terminología (fig. 3).

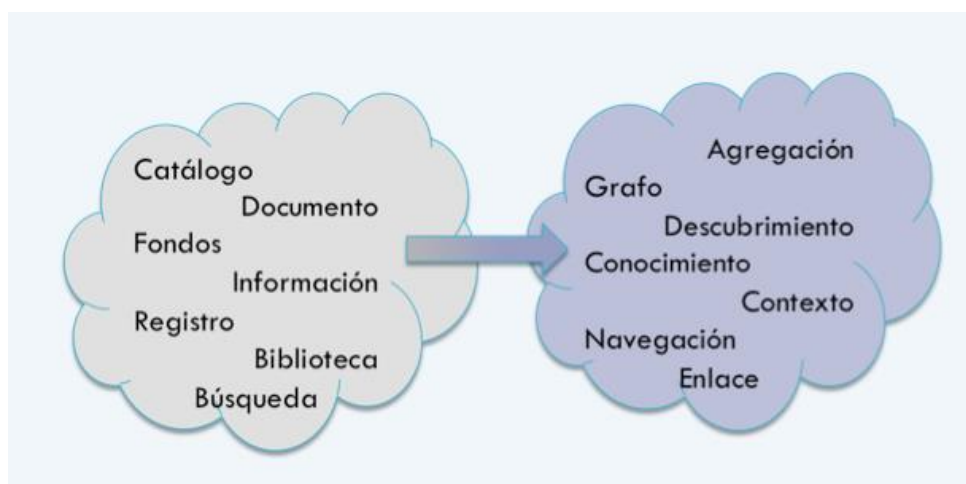


Figura 3. Cambio de terminología en las bibliotecas (trad. de Stefan Gradmann, 2012, p. 41)

5. De islas semánticas a la Linked Data Cloud

En la Linked Data Cloud (fig. 4) están recogidos hoy en día 570 datasets o conjuntos de datos, de los cuales 374 los describieron los propios proveedores en el catálogo de datasets datahub.io y 196 conjuntos de datos se identificaron a través de un rastreo automático en la Web de LOD, en abril de 2014.

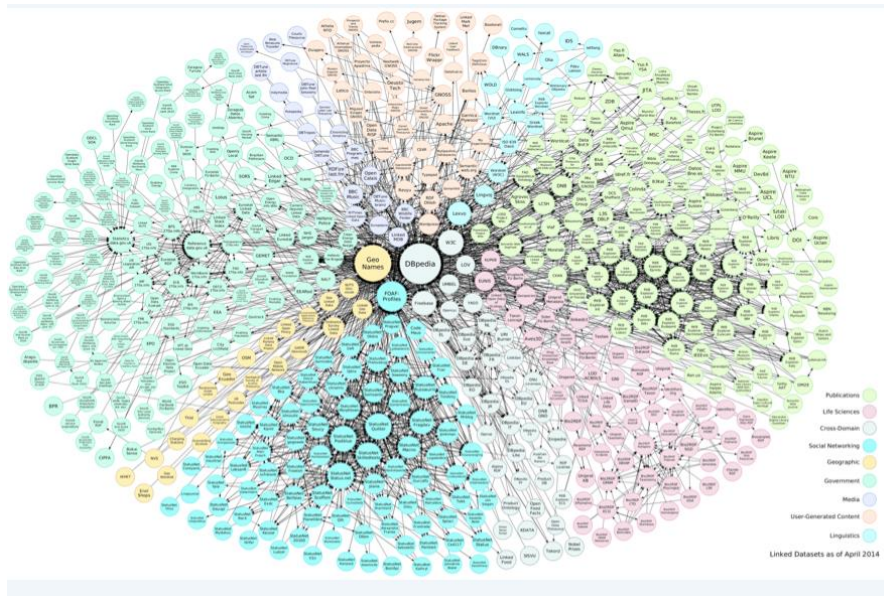


Figura 4. Diagrama de la Linked Data Cloud (Fuente: <http://data.dws.informatik.uni-mannheim.de/lodcloud/2014>)

La pregunta clave que se nos ocurre aquí, sobre todo, cuando la nube es «tan bonita», pero se ve tan poco... ¿Qué hay en esa nube de LOD que todo el mundo usa para representar el cambio en el paradigma de los datos, de las bases de datos y de todo? El diagrama de la nube LOD muestra los conjuntos de datos que se publican conforme a los principios LOD en la Web y se entrelazan con otros conjuntos de datos. Los círculos son conjuntos de datos, y los colores se corresponden con los dominios que han establecido para agruparlos: publicaciones, ciencias de la vida, redes sociales, datos geográficos, gobierno, medios, contenido generado por el usuario, lingüística, y otros independientes del dominio. Las flechas indican la existencia de enlaces entre dos conjuntos de datos RDF, y el tamaño de los círculos depende del grado de enlace de cada conjunto de datos desde otros conjuntos de datos. La nube permite identificar tanto en el centro de toda la nube (DBpedia, Geonames, etc.), como en cada dominio, los conjuntos de datos que concentran un mayor número de enlaces desde otras fuentes.

Cada vez son más las instituciones de la memoria con presencia en esta nube LOD, contribuyendo con distintos conjuntos de datos entre los que se pueden encontrar un buen número de vocabularios: tesauros como Agrovoc, encabezamientos de materia (LCSH de la Library of Congress pero también los de las bibliotecas públicas de España), clasificaciones como DDC, listas de autoridades, recursos bibliográficos, catálogos de colecciones, listas de recursos, ayudas a la búsqueda, directorios de bibliotecas, archivos y museos, entre otros.

6. Algunos casos de uso de LOD en el mundo GLAM

La mejor forma de propagar algo es por contagio. Así las tecnologías LOD se están propagando entre las instituciones de la memoria. Inicialmente han sido abrazadas por las bibliotecas (Byrne and Goddard, 2010 y Baker et al., 2012), pero también han empezado a formar parte de la Linked Data Cloud a la que nos referíamos en el apartado anterior, conjuntos de datos de archivos, museos y galerías, conformando un espacio

interesante para el desarrollo de toda la tendencia OpenGLAM. En este nuevo «reino», donde se aspira a enlazar o vincular los datos, los metadatos, es donde necesitamos un mayor entendimiento de los datos. A través de la adopción de un modelo de datos simple, se puede hacer abstracción a partir de esquemas XML y de bases de datos tradicionales, catálogos bibliográficos, etc., que se utilizaron en el pasado para interpretar y usar los datos, ahora, con la intención también de reutilizarlos.

Analizamos a continuación algunos casos emblemáticos de servicios de exposición de datos, metadatos e información en el paradigma de LOD-LAM, a los que se ha unido también recientemente (noviembre 2014) la nueva versión del catálogo de la Biblioteca Nacional de España (datos.bne.es). Lo que vaticinábamos hace quince años como una posibilidad (Méndez, 1999), se ha convertido en una realidad, y RDF es la alternativa a otras codificaciones posibles de la catalogación y la forma de exponer metadatos, no solo para mejorar la recuperación de la información, sino para abrir posibilidades a la reutilización de los datos que describen activos de información del patrimonio cultural.

6.1. Europeana: piloto de LOD del patrimonio cultural

Con muchos parabienes y otras tantas críticas, Europeana se ha convertido en el recurso, servicio o plataforma, que, por antonomasia, da acceso al patrimonio cultural europeo. La que fuera en un principio la Biblioteca Digital Europea, con el tiempo se ha convertido en «menos biblioteca y más digital» (Méndez, 2010) para potenciar el acceso al rico patrimonio europeo digitalizado, y también, cada vez más para estimular la creatividad, la innovación y la reutilización.

Una de las formas de potenciar la reutilización del contenido de Europeana es a través de Linked Open Data. El piloto de datos (en el caso de Europeana, metadatos) abiertos y enlazados data.europeana.eu, comenzó a partir de 2010 bajo los proyectos EuropeanaConnect y Europeana v2, en un esfuerzo consensuado de poner sus metadatos como LOD en la web. Esto permitiría acceder a los metadatos recopilados por Europeana, de diversos proveedores de datos, a través de tecnologías web (RDF, SPARQL, etc.), reutilizar y enriquecer esos metadatos y devolvérselos al proveedor, enriquecidos.

Los enlaces que se pueden hacer de esta forma desde los recursos de Europeana (a través de sus metadatos expuestos como LOD) con otros recursos en la Web de Linked Data que mostramos anteriormente (fig. 4) facilitarán también la recuperación de otros recursos relacionados semánticamente e igualmente abiertos y enlazados, por ejemplo, se podrían vincular dos obras de arte de un mismo artista que están relacionadas aunque no pertenezcan a la misma colección (silo) digital.

6.2. British Museum: Colección Semántica

El British Museum ha sido uno de los primeros museos en el mundo en volcar sus datos a la nube LOD. Concretamente, su catálogo disponible en línea con más de 2 millones de registros que describen 3,5 millones de objetos de museo, ha sido representado mediante la ontología CIDOC CRM (Conceptual Reference Model). Como resultado del proyecto ResearchSpace, los datos se han publicado siguiendo los fundamentos LOD: datos en el estándar RDF (notación N-Quads), base de datos de triples, punto de

consulta SPARQL 1.1, y uso de licencia de datos abiertos (British Museums' Open Data Licence 1.1).

Mediante la tecnología de Ontotext, se ofrece un interfaz de consulta más amigable que permite recuperar los objetos de la colección del museo a partir de su URI y su función como sujeto, predicado u objeto de los grafos o tripletas contenidos en su triple store. Este interfaz permite visualizar los statements en HTML y descargar en otros formatos como JSON, RDF/XML, N3/Turtle, o N-Triples.

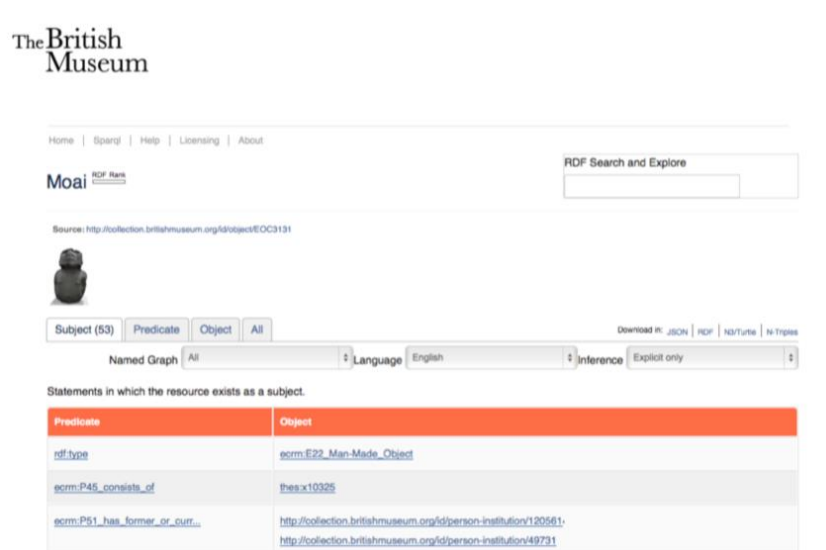


Figura 5 British Museum Semantic Web Collection (<http://collection.britishmuseum.org/resource?uri=http://collection.britishmuseum.org/id/object/EOC3131>)

6.3. VIAF como LOD: Linked Open Vocabularies

Una de las iniciativas en el reino de LOD-GLAM de mayor utilidad transversal es el fichero de autoridades internacional VIAF (Virtual International Authority File). Este proyecto iniciado en 1998, se consolidó en 2003 con la formación del Consorcio VIAF por la Library of Congress, la Deutsche Nationalbibliothek (DNB), y OCLC, a las que en 2007 se unió la Bibliothèque Nationale de France (BNF). Desde 2012 es un servicio ofrecido por OCLC y mantenido en colaboración con las bibliotecas nacionales y otros representantes de 29 países.

VIAF combina múltiples ficheros de autoridades de distintas fuentes en un único conjunto de datos enlazando las múltiples formas de autoridades (personas, geográficos, entidad, obras, expresiones) de cada una de las fuentes que lo componen. Ha probado ser altamente aplicable, pues se utiliza en un buen número de catálogos de bibliotecas, formen o no parte de la Web Semántica.

Aunque no ofrece punto de consulta SPARQL, VIAF permite la descarga completa del conjunto de datos en distintos formatos como RDF/XML, RDF N-Triples, MARC21 XML. En cuanto a su apertura, los datos están publicados bajo la licencia Open Data Commons (ODC Attribution License), que permite incluso su reutilización con fines comerciales. Pero además, el interfaz de consulta es uno de los más sencillos y a la par

amigables de los existentes en el reino LOD-GLAM, permitiendo la búsqueda y la navegación mediante diversos índices y recursos visuales.

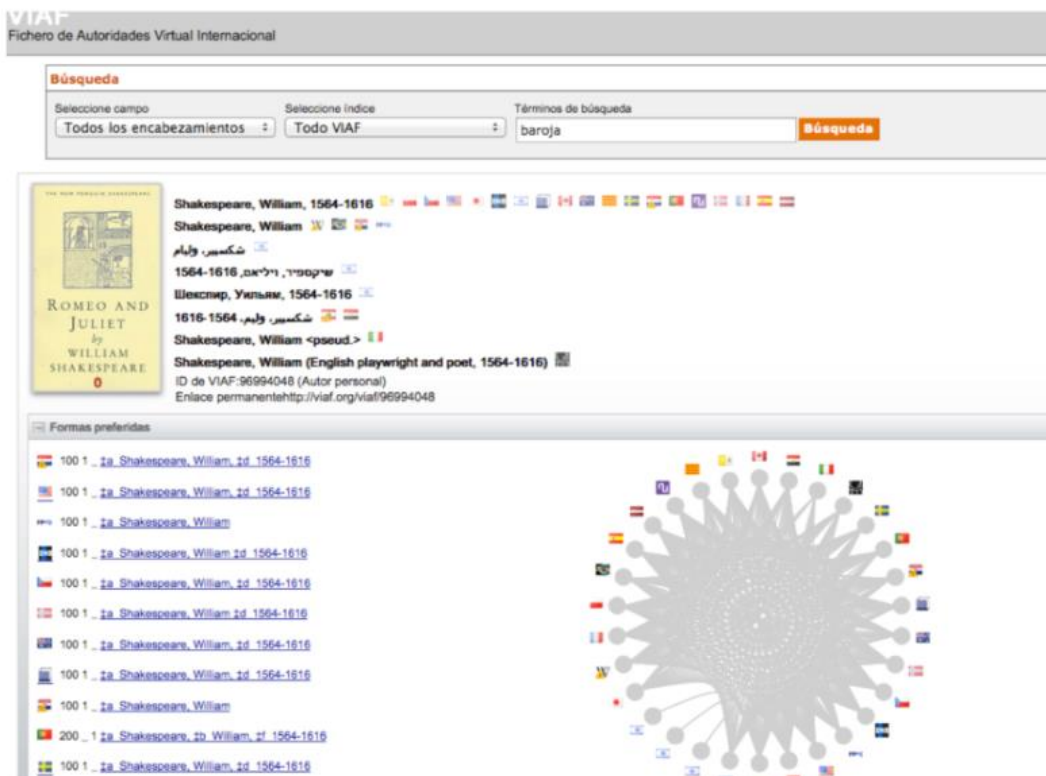


Figura 6. VIAF, Fichero de Autoridades Virtual Internacional (<http://viaf.org/viaf/96994048>)

6.4. Library of Congress: Servicio de Linked Data

El servicio de datos enlazados de la Library of Congress representa un ejemplo a seguir de la publicación de datos de bibliotecas sin tener que volcar el catálogo completo. En este caso, se ofrecen varios conjuntos de datos relativos a ficheros de autoridad (integrados a su vez en VIAF) y otros vocabularios de valores controlados, empleados en las colecciones digitales de LoC: clasificaciones, encabezamientos de materia, tesauros, listas normalizadas de países, idiomas, áreas geográficas, estándares ISO para la codificación de lenguas, y vocabularios específicos para los metadatos de preservación.

Al tratarse de vocabularios de tipo scheme, estos conjuntos de datos emplean estándares de serialización específicos como son SKOS/RDF (Simple Knowledge Organization System en RDF) y MADS/RDF (Metadata Authority Description Schema en RDF), cuyo modelo de datos ha sido mapeado a SKOS.

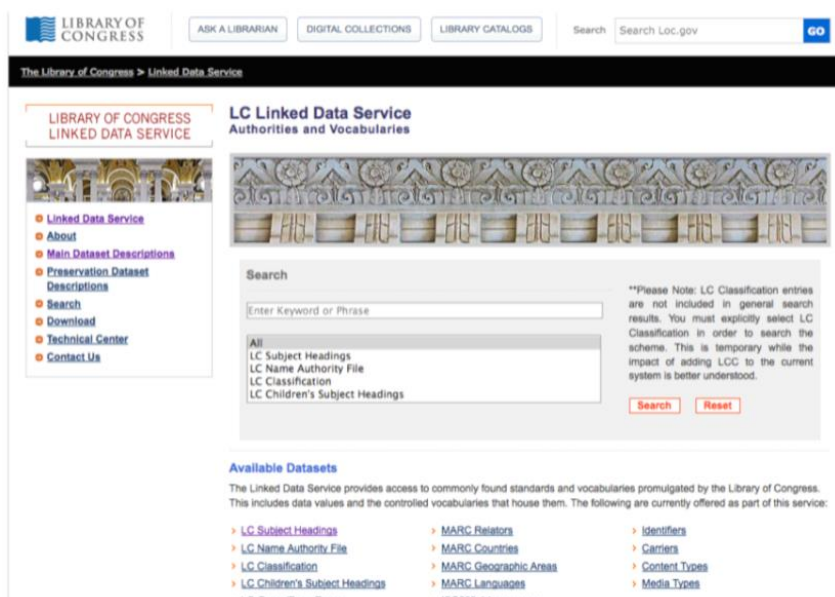


Figura 7. LC Linked Data Service: Authorities and Vocabularies (<http://id.loc.gov>)

El interfaz web de este servicio permite consultar de forma simultánea en todos los vocabularios, o filtrar por vocabularios específicos, navegar por la lista de resultados de los conceptos recuperados, visualizar la ficha completa de cada término en formato texto y en una representación gráfica, y descargar el registro correspondiente en once formatos distintos de codificación y serialización (RDF/XML, N-Triples, JSON, MADS, SKOS/RDF...). Aunque hasta el momento no dispone de punto de consulta SPARQL propio, permite la descarga en bloque de los conjuntos de datos y su reutilización en otras herramientas SPARQL, gracias al uso de una licencia de datos abierta.

7. A modo de conclusión: consejos para los GLAM en el reino de LOD

A lo largo de este artículo hemos pretendido reflejar cómo la Web Semántica es ya una realidad a través de Linked Open Data para bibliotecas, museos, archivos y galerías. Las instituciones GLAM necesitan del paradigma LOD para exhibir, o mejor, exponer los datos de sus colecciones y fomentar la reutilización de sus metadatos. Sin embargo, nos gustaría apuntar algunos consejos para los profesionales de la información a cargo de instituciones GLAM que han digitalizado o están digitalizando sus colecciones, en relación a cómo moverse en este nuevo reino de LOD-LAM:

- Decidir qué partes de los datos/metadatos son más importantes (Subject, Author, Publisher, etc.) para su publicación como LOD.
- Limpiar, enlazar y publicar los metadatos (van Holland y Verborgh, 2014) pero debemos evitar convertirnos en «el niño con el martillo» que quiere publicar todo como LOD.
- Utilizar, siempre que sea posible, identificadores persistentes para denominar las cosas (los datos, los elementos de los registros de metadatos, las palabras de los vocabularios... todos los ítems posibles que podrían reutilizarse.).

- Publicar los datos en la Web / Para la Web (RDF). Esto es, crear tantos triples como sea posible a partir de los datos, para que puedan publicarse como datos abiertos, potencialmente enlazables.
- Mantener la simplicidad: las soluciones más sencillas pueden ser suficientes para un buen resultado.

Bibliografía

*NOTA: Todas las publicaciones electrónicas se han chequeado por última vez el 10 de diciembre de 2014

BERNERS-LEE, Tim (1999): Weaving the Web: The Original Design and Ultimate

Destiny of the World Wide Web by its Inventor. San Francisco: Harper. BERNERS-LEE, Tim (2006): «Linked data». Design Issues, 2006-07-27. Disponible en: <http://www.w3.org/DesignIssues/LinkedData.html>

BERNERS-LEE, Tim (2009): «The next Web of open, linked data». TED, February. Disponible en: https://www.ted.com/talks/tim_berners_lee_on_the_next_web.html

BYRNE, Gillian and Lisa Goddard (2010): «The Strongest Link: Libraries and Linked Data». DLib Magazine, vol. 16, n. 11/12. Disponible en: <http://www.dlib.org/dlib/november10/byrne/11byrne.html>

COMISION EUROPEA (2011): Propuesta de Directiva del Parlamento y del Consejo por la que se modifica la Directiva 2003/98/CE relativa a la reutilización del sector público. Cod. 2011/0430. [Bruselas]: Comisión Europea, 2011. Disponible en: http://ec.europa.eu/information_society/policy/psi/docs/pdfs/directive_proposal/2012/es.pdf

COMISION EUROPEA (2014): Comunicación de la Comisión al Parlamento europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las regiones. Hacia una economía de los datos próspera. COM(2014) 442 final. [Bruselas]: Comisión Europea, 2014. Disponible en: http://datos.gob.es/sites/default/files/com_2014_442_f1_communication_from_commission_to_institutes_v2_p1_773238_4.pdf

GRADMANN, Stefan (2012). «From Containers to Content to Context: the Changing Role of Libraries in eScience and eScholarship». Bielefeld Conference, April 24 2012. Disponible en: http://conference.ub.uni-bielefeld.de/programme/presentations/Gradmann_BC2012.pdf

HEATH, Tom and Christian Bizer (2011): Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool. Disponible en: <http://www.morganclaypool.com/doi/abs/10.2200/S00334ED1V01Y201102WBE001>

HEATH, Tom, and Christian Bizer (2011): Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool. Disponible en: <http://linkeddatabook.com/editions/1.0/>

MENDEZ, Eva (1999): «RDF un modelo de metadatos flexible para las bibliotecas del próximo milenio». Jornadas Catalanas de Documentación (1999. 7a. Barcelona). Disponible en: <http://www.cobdc.org/jornades/7JCD/1.pdf>

MENDEZ, EVA (2004): «La Web Semántica: una web más bibliotecaria». CLIP: Boletín de Sedic, n. 41, 2004. Disponible en: http://www.sedic.es/p_boletinclip41_confirma.htm

MENDEZ, Eva and Jane Greenberg (2012): «Datos enlazados para vocabularios abiertos: marco global de HIVE». El Profesional de la Información, 21 (3): 236-244. Disponible en: http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/03_esp.pdf

MINISTERIO DE LA PRESIDENCIA (España) (2011): Real Decreto 1495/2011, de 24 de octubre, por el que se desarrolla la Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público, para el ámbito del sector público estatal. Boletín Oficial del Estado, núm. 269 de 8 de noviembre de 2011, pp. 116296-116307. Disponible en: http://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-17560

PARLAMENTO Europeo; Consejo (2013): Directiva 2013/37/UE del Parlamento Europeo y del Consejo de 26 de junio de 2013 por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público. Diario Oficial de la Unión Europea. Disponible en: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:ES:PDF>

PROYECTO APORTA (2009): Guía Aporta sobre reutilización de la información del sector público. Madrid: Secretaria de Estado de Telecomunicaciones y para la Sociedad de la Información. Disponible en: http://www.osimga.org/export/sites/osimga/gl/files/guia_reutilizacion_3M.pdf

VAN HOLLAND, Seth, and Ruben Verborgh (2014): Linked Data for Libraries, Archives and Museums. How to clean, link and publish your metadata, London: Facet.

W3C Incubator Group (2012): Library Linked Data Incubator Group Final Report. Thomas Baker et al. Disponible en: <http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025> (En español: <http://skos.um.es/Incubator/lld/XGR-lld-vocabdataset>)