

J.12. Cómo *linked open data* impactará en las bibliotecas a través de la innovación abierta

Tomás Saorín

5 diciembre 2011

Saorín, Tomás. "Cómo *linked open data* impactará en las bibliotecas a través de la innovación abierta". *Anuario ThinkEPI*, 2012, v. 6, pp. 288-292.



Resumen: Se aborda la importancia del estímulo de la innovación abierta a partir de los recursos de información de bibliotecas, archivos e instituciones culturales. La Web se está convirtiendo en un espacio global de datos, en el que las bibliotecas han de participar publicando sus datos usando tecnologías RDF y licencias abiertas, para potenciar servicios innovadores desarrollados por terceros, reutilizando sus datos.

Palabras clave: Datos enlazados abiertos, Innovación abierta, Bibliotecas, Reutilización de información.

Title: Open innovation for libraries using linked open data technologies

Abstract: The importance for libraries, archives and cultural heritage institutions of open innovation on the basis of linked open data basis is discussed. As the web evolves into a global data space, libraries should provide their datasets using RDF

technologies and open licenses to strengthen innovative services developed by third parties reusing their metadata and digital contents.

Keywords: Linked open data, Open innovation, Libraries, Information reuse.

Linked data es una apuesta tecnológica segura, y conviene comprender a qué sectores de la información afecta. No es únicamente un importante desarrollo tecnológico, sino que también tiene un componente de innovación en el mercado de la información. Se pretende aquí precisar su propuesta de valor y dar más razones para aprovecharla.

La innovación abierta puede multiplicar la oferta de servicios bibliotecarios

Muchas notas *ThinkEPI* han tratado el tema de la innovación. **Lara-Navarra** (2010) ofrecía una síntesis de los modelos de innovación existentes, y de la importancia de la sistematización de la innovación en la práctica de la profesión. Existe abundante bibliografía sobre el tema, es un eje básico de cualquier estrategia de desarrollo empresarial y competitividad (**Escorsa-Castells; Valls-Pasola**, 2003).

Aunque las bibliotecas han estado activas en la evolución de internet, especialmente en sus primeras etapas, sería fantástico creer que siguen ocupando en la actualidad un lugar central en el mundo digital en contenidos digitales, interac-

ción, búsqueda, relevancia social, liderazgo de opinión y, por supuesto, innovación. Aunque se han planificado y desarrollado servicios digitales innovadores, su impacto es muy limitado y su propuesta queda lejos de otras iniciativas. Su innovación es de corto alcance. Las bibliotecas no han hecho *Wikipedia* ni *Facebook*.

Entonces, *linked open data* ¿es para las bibliotecas una amenaza o una oportunidad?, ¿conviene esperar o es el momento de incorporarse? Es momento de acordar cuál es la posición en la que las bibliotecas pueden ser más útiles y eficientes.

Linked data es una transición necesaria para las bibliotecas y proyectos de información culturales. *Europeana* estaría señalando el camino a seguir, así como los esfuerzos por conseguir gestionar una transición bibliográfica desde la *Library of Congress* (2011). El reciente informe sobre *Datos enlazados bibliotecarios* del *Grupo Incubadora del W3C* (LLD XG, 2011), es relevante para entender la oportunidad que se abre para las bibliotecas de hacer mejor las cosas, alcanzar más impacto digital y apostar por formas complementarias de utilidad social indirecta. Ya disponemos de bibliografía de referencia para trazar un mapa de iniciativas y elementos básicos del campo *linked open data* (**Stuart**, 2011; **Peset; Ferrer-Sapena; Subirats-Coll**, 2011).

Linked data es un mundo complejo, y por lo tanto conviene seleccionar en qué se van a centrar nuestros recursos. En las siguientes líneas se argumenta a favor de centrarse en la publicación de *datasets* (conjuntos o fuentes de datos) para permitir la innovación abierta.

Innovación abierta

Término acuñado por **Henry Chesbrough**, que hace referencia a la apertura del proceso de investigación de una empresa a terceros. Supone delegar en agentes externos a la organización parte de las tareas de desarrollo de innovaciones, compartiendo con ellos nuestros datos y nuestra información (**Chesbrough**, 2006).

El autor defiende que, en un mundo caracterizado por el conocimiento distribuido, las empresas pueden crear más valor y explotar mejor sus procesos de investigación si integran conocimientos, objetos de propiedad intelectual y productos externos en su trabajo de innovación. El término “capilaridad” (**Cornella; Flores**, 2007) se refiere a la capacidad de comunicación intensificada entre todos los niveles internos y externos de la organización para ampliar las posibilidades de innovación, y pone énfasis en la información que se intercambia con los clientes, proveedores, socios y subcontratistas.

Incorporar el entorno es sólo una de las caras de la innovación abierta. La otra es dejar que sea alguien de fuera el que la desarrolle de forma autónoma. Se conoce como *inside-out*, innovación de dentro-afuera, a la que genera rentabilidad de activos que nuestra organización no es capaz de poner en valor. La naturaleza de las tecnologías *linked data* se adapta perfectamente a esta forma de innovación, cada vez más frecuente en otros campos: dejar que otros busquen beneficio con algo que apenas sabemos qué hacer.

Uno de los patrones para modelos de negocio (**Osterwalder; Pigneur**, 2011) que tiene especial incidencia en la información digital es el de rentabilizar productos, tecnologías, conocimientos y objetos de propiedad intelectual que no se utilizan en la empresa poniéndolos a disposición de terceros.

Las instituciones públicas (y entre ellas las bibliotecas) suelen no estar en condiciones de obtener todo el valor de sus activos en información por limitaciones conceptuales y por falta de inversiones para la innovación.

Las políticas públicas de reutilización de información están tratando de crear conciencia del valor económico de liberar los datos, porque son una de las nuevas materias primas de la economía de la información. Los datos públicos pueden además liberarse de la forma más abierta posible,

de modo que estaríamos refiriéndonos a “innovación abierta x2”, porque es abierta a que otros le busquen utilidad y basada en datos abiertos (*open data & open innovation*).

Publicar primero; consumir, ya veremos

La dualidad más evidente ante los datos enlazados es el equilibrio entre publicar y consumir. Publicar supone adaptar nuestros sistemas para generar fuentes de datos de calidad, basadas en vocabularios RDF y enlazadas a recursos de otros *datasets*. Consumir supone la explotación de los datos externos disponibles para aportar valor a nuestros propios servicios digitales: combinar datos en una aplicación para el usuario final.

Las bibliotecas han de apostar por publicar con calidad transformando los sistemas de catalogación y de acceso público y difundir sus *datasets* entre los agentes infomediarios, no entre los usuarios finales. Hay que comprender los patrones de publicación de datos enlazados, los procesos para asegurar su calidad y la transformación desde diferentes modelos de sistemas de almacenamiento de contenidos. El capítulo 5 del libro online *Linked data: evolving the Web into a global data space*, de **Heath y Bizer** (2011) es una buena lectura práctica para publicar *linked data*.

El término *infomediarios* es central en la cadena de valor propuesta en el *Proyecto Agenda* para la reutilización de datos públicos (**Onsti**, 2011). Nuestro público, en el contexto de los datos enlazados, son los agentes de todo tipo que están en disposición de concebir, explotar y poner en práctica propuestas innovadoras, y que ya disponen de tecnologías y pueden encontrar líneas de negocio.

A través de ellos se llegará al usuario final, de forma que los datos bibliotecarios serán en cierto modo invisibles al consumidor. Ya hemos sugerido alguna vez la oportunidad de algo así como una etiqueta “*library data inside*” (una forma sencilla de expresar “para hacer el producto que está usted usando se han utilizado datos procedentes de la red de bibliotecas públicas...”) para obtener beneficios de posicionamiento de la imagen del servicio de bibliotecas.

En este contexto de desintermediación hay que idear también entre todas formas de hacer visible la participación de las bibliotecas en lugares a los que difícilmente habríamos llegado por nosotros mismos.

También las bibliotecas, especialmente bajo la forma de proyectos de redes e intersectoriales, pueden explorar la capacidad de combinar datos culturales, impulsando iniciativas de consumo de

datos enlazados. Aunque esta es una de las líneas de *Discovery, a metadata ecology for UK education & research*, no debe entenderse en términos de exclusividad, como por ejemplo los catálogos actuales de bibliotecas, sino otra propuesta más dentro de la oferta en internet. Si un indicador frecuente de éxito es el número de visitas a una web, en *linked data* se obtiene mayor valor cuantas más veces sea integrada la información de las bibliotecas en productos de información de otras entidades.

¿Qué tienen las tecnologías LOD para ser tan apropiadas para la innovación abierta?

La integración de datos se hace más fácil. Se reducen las barreras para que los servicios incorporen masivamente datos. Las actuales APIs que permiten realizar *mashups* exigen un esfuerzo individualizado para cada una de ellas, lo que limita su uso o focaliza la capacidad de aplicación a unas pocas fuentes. El reciente anuncio de *Spotify* para que se construyan aplicaciones sobre su API implica que los desarrolladores tienen que conocer un nuevo lenguaje, que sólo podrán rentabilizar en una única fuente.

“Una de las actividades de la biblioteca *linked data* será hacer que se conozcan bien sus fuentes de datos y promocionar su uso”

RDF y *Sparql* suponen la creación de un mercado amplio para los datos en el que las empresas puedan competir por ofrecer servicios diferenciados, en lugar de por islas de datos que aspiran a copar un mercado. Que exista un lenguaje universal para la interrogación de fuentes de datos de cualquier naturaleza es un avance de enorme valor. Estamos poniendo piezas sólidas para un “único espacio de datos global” (Heath; Bizer, 2011).

La arquitectura del grafo de datos enlazados permite descubrir fuentes de datos, explorando las relaciones que enriquecen los datos publicados. Una de las actividades de la biblioteca *linked data* será que se conozcan bien sus fuentes de datos y promocionar su uso entre tantas opciones como pueda haber. El slogan AAA de la web semántica “*anyone can say anything about any topic*” (Allemang; Hendler, 2008). Esto fomenta la pluralidad de la información, permitiendo que muchos agentes agreguen información sobre las mismas entidades para hacer diferentes puntos

de vista simultáneos. Para estos autores, cada vez cobra más sentido la asunción de *open-world*: los datos no pueden concebirse como completos o acabados, siempre cabe la posibilidad de ampliación para una entidad dada.

Son piezas de un discurso abierto, el cual se soporta en la naturaleza distribuida e interconectada de RDF. Claro que, como en todo espacio abierto, aparecerá el *data spam*, la manipulación de datos y los intentos de explotación ilegítima. Por otro lado, cierta redundancia en los datos disponibles ayuda también a optimizar los sistemas intensivos en uso de datos distribuidos, y a reducir las fases de puesta en marcha de servicios.

Los vocabularios y esquemas de metadatos son públicos y en línea, de forma que el negocio bibliotecario, hasta ahora muy exclusivo con sus formatos y modelos de datos, pasa a ser global: pueden ser incorporados algunos elementos bibliotecarios a un proyecto sobre cine o cocina, o viceversa, las bibliotecas pueden enriquecer sus datos con fuentes de origen sectorial hasta ahora inaccesibles. Ya no existen modelos puros, todo puede ser recombinado, y cuantos más elementos comunes se usen, más abierto se está a la reutilización en la Web.

Esto también tiene importantes implicaciones en la irrupción de nuevas empresas en el mercado de soluciones informáticas para bibliotecas, hasta ahora un mercado dominado por grandes y caros productos. La descentralización también tiene implicaciones económicas, ya que los costes de puesta en marcha, escalabilidad y despliegue se comparten entre proveedores de datos y consumidores, y permite avanzar e invertir (*pay-as-you-go*). Que las inversiones puedan distribuirse entre agentes y en el tiempo es una clave para su éxito en la economía real.

“Forse altro canterà con miglior plectro”

Este verso de Ariosto, que usó Cervantes al cerrar el primer Quijote, nos sirve para insistir en que, en un espacio digital hiperconectado, facilitar la reutilización podrá provocar que otros produzcan innovación a partir de nuestra oferta de datos. “Quizá otro tendrá más inspiración para cantar” significa primero, que desde otros puntos se añadirá información sobre el mismo ítem, y también que es seguro que otros tendrán buenas ideas, conocimientos, medios y una visión que aportará valor a los datos procedentes de bibliotecas.

Es frecuente que la mejor propuesta de valor esté fuera de la propia institución. Entendiendo como tal aquella que conecta con una necesidad de un segmento de usuarios y que se despliega

ción de modelos de negocio. Ediciones Deusto, 2011. ISBN: 978 8423427994

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, v. 10, n. 2, pp. 165-173.

<http://www.elprofesionalde lainformacion.com/contenidos/2011/marzo/06.pdf>

<http://dx.doi.org/10.3145/epi.2011.mar.06>

Stuart, David. *Facilitating access to the Web of data: a guide for librarians*. London: Facet Pub., 2011. ISBN: 978 1856047456

* * *

Linked open data y Europeana data model

Xavier Agenjo



Las actividades claves de *Europeana* son agregar, facilitar, distribuir y atraer. Es decir, abrir datos, vincular datos y publicarlos, y construir aplicaciones que atraigan a diferentes usuarios. Entre los detalles de los *Key performance indicators* para 2012: buscar la

participación de editores, ampliar el contenido a otros medios como los materiales audiovisuales e implementar completamente EDM en el flujo de trabajo de ingesta de datos. Y todo ello afecta a los LAM que participan en *Europeana* en un camino de ida y vuelta, abrir datos y reutilizar.

Es también previsible que se tomen importantes acuerdos acerca de la estructura organizativa de *Europeana* y, sobre todo, se proceda a hacer oficial el *Europeana data exchange agreement*¹ que, entre otras muchas cosas, aboga decididamente por el *Creative Commons Universal Public Domain Dedication*², y, lo que es más importante, incluye la autorización para la reutilización de los contenidos agregados por *Europeana* a cargo de terceras partes.

El estado de la cuestión en lo que se refiere al *Europeana data model* se ha visto consolidado con la edición de una serie de documentos: el *Europeana data model primer*³, que actúa claramente a modo de directrices. El documento apareció el 26 de octubre de 2011. Al día siguiente apareció publicado el *Europeana data model mapping guidelines*⁴, en el que se describen minuciosamente las directrices para transformar los documentos a

los *namespaces* y clases -tanto básicas como contextuales- de EDM mediante un amplio conjunto de tablas y propiedades. El mismo día 27 de octubre apareció publicado un documento análogo a los dos primeros, el *Europeana data model fact sheet*⁵, mucho más resumido y dirigido a quienes deben tomar decisiones basándose en premisas técnicas.

Europeana está ejerciendo un importantísimo liderazgo en este sentido, sus especificaciones funcionales, su modelo de datos y su normativa está contribuyendo a delimitar un camino que muchos LAM encuentran más efectivo que andar en solitario. *Linked open data* tiene un efecto fundamental en archivos, bibliotecas y museos que no es otro que el de ampliar el ámbito de aplicación de los datos que generan los LAM y, sobre todo, el de ampliar los puntos de vista y análisis de la información.

LOD es, en definitiva, abrir datos, enlazar datos, publicar datos enlazados; el papel de los LAM es aportar y reutilizar.

La *Fundación Ignacio Larramendi* anunció el pasado día 17 de noviembre la *Biblioteca Virtual de la Escuela de Salamanca*⁶.

Se está aplicando el modelo de *linked open data* tanto a los autores como a las obras de esos autores que constituyen la *Escuela de Salamanca*. Se trata de editar en abierto, es decir en *open access*, esos objetos digitales, y se trata también de vincular esos objetos digitales con otros, utilizando una tecnología concreta, *linked open data*.

Notas

1. http://www.version1.europeana.eu/c/document_library/get_file?uuid=deb216a5-24a9-4259-9d7c-b76262e4ce55&groupId=10602
2. <http://creativecommons.org/publicdomain/zero/1.0>
3. http://version1.europeana.eu/c/document_library/get_file?uuid=48b552e1-c71d-4f3f-a1ae-c1929cb7de76&groupId=10605
4. http://version1.europeana.eu/c/document_library/get_file?uuid=99ce6a74-8e55-4321-917a-65bdf1fe5bc&groupId=51031
5. http://version1.europeana.eu/c/document_library/get_file?uuid=ef2baffc-f078-41d9-be5f-76a3427f198f&groupId=51031
6. http://www.larramendi.es/i18n/cms/elemento.cmd?id=estaticos/paginas/Biblioteca_Virtual_Escuela_de_Salam.html

agenjo@ono.com