DATOS ABIERTOS Y REPOSITORIOS DE DATOS: NUEVO RETO PARA LOS BIBLIOTECARIOS

Tony Hernández-Pérez y **María-Antonia García-Moreno**

Tony Hernández-Pérez es doctor en ciencias de la información y profesor del *Depto. de Biblioteconomía y Documentación* de la *Universidad Carlos III de Madrid* en donde actualmente dirige el programa de doctorado en documentación. Su labor docente e investigadora está ligada al grupo TecnoDoc incluyendo asignaturas de web social, gestión de contenidos web, metadatos, búsqueda y recuperación de información, e-learning y documentación periodística y audiovisual.
http://orcid.org/0000-0001-8404-9247

Univ. Carlos III de Madrid, Depto. de Biblioteconomía y Documentación
C/Madrid, 126. 28903 Getafe (Madrid), España
tony@bib.uc3m.es

María-Antonia García-Moreno es doctora en ciencias de la información y profesora del *Depto. de Biblioteconomía y Documentación* de la *Universidad Complutense de Madrid*. Directora del grupo de investigación *Bisoc*. Imparte docencia sobre tecnologías de la información –bases de datos y automatización-, así como en metodología de la investigación.
http://orcid.org/0000-0002-2369-5488

Univ. Complutense de Madrid, Fac. de Ciencias de la Documentación
Santísima Trinidad, 37. 28010 Madrid, España
mariaant@ccinf.ucm.es

Resumen

Se analiza el concepto de datos abiertos, que debe cumplir muchos más requisitos que el de estar simplemente accesibles a través de internet en cualquier formato. Se ofrecen algunos ejemplos sobre la importancia de los repositorios de datos científicos en abierto para favorecer la transparencia en el desarrollo de la ciencia y evitar el fraude científico, y se ponen de manifiesto algunos de los problemas en la gestión de estos repositorios de datos.

Palabras clave

Acceso abierto, Datos abiertos, Datos científicos, Datos de investigación, Repositorios.

Title: Open data and data repositories: a new challenge for librarians

Abstract

The concept of open data, which demands that many more conditions be met than simple internet accessability in any format, is analyzed. Some examples are given about the importance of open data repositories in efforts to promote transparency in science and avoid scientific fraud. Some of the main problems in the management of data repositories are discussed.

Keywords

Open access, Open data, Scientific data, Research data, Data sharing, Repositories.

Hernández-Pérez, Tony; **García-Moreno, María-Antonia** (2013). "Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios". *El profesional de la información*, 2013, mayo-junio, v. 22, n. 3, pp. 259-263.
http://dx.doi.org/10.3145/epi.2013.may.10

Introducción

En los últimos años los artículos, noticias y sopas de siglas relacionados con datos no han hecho más que crecer: *big data* (*Schumpeter*, 2011), *open data* (*Open Knowledge Foundation*, 2012), *linked data* (**Berners-Lee**, 2009), (**Ríos-Hilario**; **Martín-Campo**; **Ferreras-Fernández**, 2012), *sharing data* (**Torres-Salinas**; **Robinson-García**; **Cabezas-Clavijo**, 2012), *open linked data* (**Méndez**; **Greenberg**, 2012). En general, se habla de datos estadísticos, geográficos, de transporte, meteorológicos, financieros, medioambientales, gubernamentales, científicos y culturales. Las tecnologías de la información están permitiendo recopilar grandes cantidades de datos:

- Datos personales, nuestros hábitos de compra (tarjetas de crédito), sobre nuestras relaciones personales y/o profesionales (*Twitter* o *Facebook*), sobre nuestros gustos (los "me gusta" o "compartir" de muchas redes o las visualizaciones de vídeo en *YouTube*), nuestros viajes, o nuestra salud; lo que corremos, lo que andamos o las calorías que ingerimos o que gastamos mediante aplicaciones que

existen en nuestros móviles, muy fáciles de activar.
- Datos sobre el contexto en el que vivimos, sobre nuestro medio ambiente, desde los niveles de contaminación o cantidad y tipos de partículas de polen en el aire hasta las viviendas en venta o en alquiler en la calle en que me encuentro en un momento dado.
- Datos sobre objetos y productos, desde conocer la historia de un edificio a partir de una foto tomada con el móvil hasta la procedencia y la fecha de envasado y caducidad de cualquier producto alimenticio, o los datos de edición de cualquier libro a partir de lectores de códigos de barras.
- Datos sobre los procesos, lo que tarda un autobús en llegar a una parada, la ruta en tiempo casi real de un avión comercial, la localización, carga, origen y destino de un barco o el tiempo que pasamos leyendo una página web.

Existe cierta creencia de que los datos, por el simple hecho de que estén disponibles para ser leídos o descargados en la Web, por ejemplo en formato pdf, son datos abiertos. No, no lo son o lo son en un grado muy restringido. No todos los datos que se recopilan o todos los datos que se pueden descargar de la web son datos abiertos. Para que sean realmente abiertos sí deben estar disponibles en internet, preferiblemente para ser descargados, pero también deben poseer algún tipo de licencia legal para poderlos utilizar, reutilizar y redistribuir, mezclándolos incluso con otros datos, sujetos como mínimo a la "atribución" (reconocimiento de la autoría, quién lo ha hecho), o al "compartir igual" (que la explotación que se haga de esos datos –incluyendo las obras derivadas- mantengan la misma licencia al ser divulgadas).

La definición de "abierto" se entiende dentro del contexto que aprueba el *Open Definition Advisory Council*, una organización con instituciones y representantes personales destacados del movimiento de acceso abierto en donde queda claro que el término conocimiento incluye contenidos, como música, películas o libros pero también "datos tanto científicos, históricos, geográficos o de cualquier otro tipo e información gubernamental y de otras administraciones públicas"[1]. La definición recoge 11 condiciones para que una obra o unos datos sean considerados abiertos: acceso (disponible integralmente, a un coste razonable y de forma que pueda ser modificable), redistribución, reutilización, ausencia de restricciones tecnológicas, reconocimiento, integridad, sin discriminación de personas o grupos, sin discriminación de ámbitos de trabajo, distribución de la licencia, la licencia no debe ser específica de un paquete y la licencia no debe restringir la distribución de otras obras.

Tim Berners-Lee, el inventor de la Web pero también de la iniciativa de la web semántica y los datos enlazados (*linked data*) ha propuesto un esquema de cinco estrellas para los datos abiertos[2]: una estrella, para poner los datos en la Web, en cualquier formato, con una licencia abierta. Por ejemplo, un documento en pdf. Este tipo de documentos permite leer información pero puede resultar complejo extraer información de ella: aún se recuerda en España la polémica decisión del *Congreso* y del *Senado* de publicar la declaración de bienes de sus miembros en formato pdf[3].

Las dos estrellas del esquema de **Berners-Lee** se otorgan cuando además los datos se ofrecen en un formato estructurado, por ejemplo, en *Excel* en vez de imágenes o pdf. Las tres estrellas, cuando en vez de en un formato propietario (por ejemplo, *Excel* pertenece a *Microsoft*), los datos se ofrecen en un formato no propietario: por ejemplo, ficheros de datos separados por comas (*comma separated values*, csv) para que puedan ser tratados en el programa que cada uno prefiera. Cuatro estrellas, cuando esos datos utilizan uris (*uniform resource identifiers*), por ejemplo urls, para poder identificar cosas y para que otros puedan apuntar hacia ellas. Y cinco estrellas cuando se enlazan a otros datos para proporcionar contexto.

Importancia de los datos científicos

En el ámbito de los datos abiertos fundamentalmente se distingue entre datos científicos y datos gubernamentales. Como es lógico muchos datos científicos han sido financiados a través de agencias públicas por lo que también pueden ser considerados gubernamentales. No obstante, nos referimos a estos últimos cuando hablamos de datos que emanan no de centros de investigación sino de las administraciones públicas (datos de transporte, geográficos, económicos, culturales, etc.).

Cuando se habla de la ventaja de los datos abiertos a menudo se mencionan tres tipos de beneficios: la transparencia para el buen funcionamiento de las sociedades democráticas puesto que permite conocer lo que hacen los gobiernos; el aporte del valor comercial y social puesto que permite la creación de negocios y servicios innovadores basados en esos datos; y la participación y el compromiso de los ciudadanos puesto que al estar más informados pueden implicarse más en los procesos de tomas de decisiones.

En este artículo nos centraremos en la importancia que tienen para la biblioteca los datos abiertos en el entorno científico. Y en la importancia y las políticas para impulsar este tipo de repositorios en las organizaciones dedicadas a la investigación. La transparencia abarca también el ámbito de la ciencia, incluida las ciencias sociales. En abril de 2013 tuvimos un ejemplo que se puede analizar a partir del artículo sobre "la depresión del Excel" de **Paul Krugman** (2013) en el que se explicaba cómo, a partir del análisis de los datos originales de **Reinhart** y **Rogoff**, –los datos científicos originales utilizados para su investigación- algunos investigadores de la *University of Massachusetts* demostraron la existencia de errores que podrían poner en entredicho los resultados de su investigación. **Reinhart** y **Rogoff** son los economistas a los que se les atribuye las bases para la decisión de las políticas económicas de austeridad que imperan al menos en Europa y cuyos argumentos se encuentran en el artículo cuestionado.

Los editores de revistas científicas comienzan a mostrar su preocupación por las malas conductas de algunos investigadores, quienes presionados por la necesidad de publicar en revistas con factor de impacto están aumentando de forma alarmante los casos de fraude científico a través del plagio, autoplagio o el fenómeno de citación coercitiva (**Martin**, 2013). A esto hay que añadir el que algunos científicos falsifican los datos (**Fanelli**, 2009) que supuestamente demuestran sus hipótesis, como el famoso caso de las células madre del científico coreano Hwang Woo-Suk. Por tanto, cuando

se habla de la transparencia como una de las ventajas de los datos abiertos, no sólo nos referimos a la transparencia en el sentido socio-político del término sino también a la transparencia científica.

Repositorios de datos científicos abiertos

¿Y qué tiene esto que ver con los bibliotecarios o con los documentalistas? Uno de los muchos sectores que más datos recopila es el científico, en gran parte financiado con fondos públicos. Datos climáticos, geográficos, de vida marina, astronómicos o económicos. La mayor parte de ellos sirven de apoyo a la publicación de artículos que son publicados mayormente en revistas científicas de los que, con suerte, hasta un 25% de media a escala mundial pasan a formar parte de los repositorios institucionales (**Ginsparg**, 2011; **Björk** *et al.*, 2010; **Gargouri** *et al.*, 2012).

La mayor parte de las universidades y centros de investigación de todo el mundo, incluido nuestro país, disponen ya de repositorios institucionales que almacenan los resultados de investigación de sus miembros (principalmente artículos, comunicaciones a congresos y tesis doctorales). Pero además, en un futuro próximo será imprescindible, tal y como recomienda la *OCDE*, *Organisation for Economic Cooperation and Development* (2007)[4], la *Comisión Europea* (*eIRG eInfrastructure Reflection Group*, 2009)[5] y obliga ahora la *NSF*, *National Science Foundation* (2011)[6], que todos los proyectos que soliciten financiación deberán presentar un plan de gestión de datos de investigación con el fin de que puedan ser compartidos, como explicaba **Daniel Torres-Salinas** en *ThinkEPI* (2009) o como nos contaba **Jane Greenberg**[7] en su conferencia "Bibliotecas digitales para datos de investigación"[7] en donde relató su experiencia en *Dryad*, uno de los repositorios de datos más importantes en el campo de la biología, la ecología y la medicina.

> A medio plazo, las universidades y centros de investigación deberán poner en marcha un repositorio de datos de investigación o realizar acuerdos para colaborar con alguno

Por tanto, previsiblemente a medio plazo, las universidades y centros de investigación, además de contar con el repositorio institucional que ya está funcionando en la mayoría de estos organismos, deberán poner en marcha un repositorio de datos de investigación o realizar acuerdos en el que colaborar con alguno (temático o interinstitucional). El hecho de que sean repositorios independientes no significa que no exista relación entre ellos. Los repositorios de datos de investigación sirven, entre otros fines, para validar resultados de investigación y, por tanto, deben estar vinculados de alguna manera a las publicaciones científicas en donde se muestra para qué fueron utilizados esos datos, por lo que algunos de los problemas se podrán abordar de formar conjunta, tanto para los repositorios institucionales como para los repositorios de datos de investigación.

Partimos pues de la hipótesis de que deben ser repositorios distintos a los repositorios institucionales, puesto que las características de ambos difieren sustancialmente, tanto respecto a los objetivos para los que sirven uno y otro, como a las técnicas de gestión y mantenimiento, las políticas de acceso y depósito o la tipología de datos y cantidades de datos a preservar, así como la importancia de la procedencia y la validez de los datos, como ya sostienen algunos autores (**Arano-Poggi** *et al.*, 2011). Justamente, porque el perfil del bibliotecario cuenta con la confianza, el espíritu de interdisciplinariedad y colaboración y la experiencia en la gestión de datos digitales y en su preservación es por lo que entendemos debe ser desde la biblioteca desde donde se aborden los proyectos de repositorios de datos científicos.

Más allá de la infraestructura tecnológica los planes de gestión de datos deben contemplar:

- la descripción de los datos;
- la definición de los estándares de calidad;
- propiedad intelectual y derechos;
- licencias; y
- políticas de archivos y preservación.

No será una tarea fácil: existen muchos tipos de datos, desde los que se usan en cualquier tipo de ingeniería a los de un arqueólogo. Muchas veces los datos se encuentran dispersos y aún no existe ni siquiera conciencia de la necesidad de hacer copias de seguridad, y la cantidad de almacenamiento necesaria supera generalmente las capacidades que proporciona la universidad.

Hay dos aspectos que posiblemente sean clave:

1) muchos investigadores son reticentes a compartir unos datos que les ha costado mucho esfuerzo obtener para compartirlos fuera; y

2) la gestión de datos requerirá un reciclaje tanto de los profesionales como de los planes de formación para superar posibles lagunas relacionadas con el conocimiento y las habilidades en el manejo de grandes cantidades de datos generalmente numéricos.

Para ayudar a superar estos inconvenientes en la puesta en marcha de repositorios y de planes de datos para los científicos –algunos obligados a presentarlos para obtener financiación-, algunas universidades anglosajonas y el inglés *Digital Curation Centre* han elaborado informes para ayudar a concienciar y realizar un plan de gestión de datos (*data management plan*, DMP)[8].

Los catálogos de repositorios más importantes probablemente sean *Databib*, con más de 500, y el *re3data.org Registry* (**Pampel** *et al.*, 2013). *Databib* no diferencia y recoge todos los modelos de repositorios: temático basado en el trabajo conjunto con los editores, como *Dryad*; de consorcios, como el holandés de *Datacentrum*; nacional, como el australiano *Research Data Australia*; o institucional, como el de la *Purdue University*, o *Ideals* de la *University of Illinois*, cuyo modelo integra el repositorio de datos dentro del repositorio institucional de resultados de la investigación, en nuestra opinión una opción menos eficiente.
http://databib.org
http://datadryad.org
http://datacentrum.3tu.nl
http://researchdata.ands.org.au

https://purr.purdue.edu
https://www.ideals.illinois.edu

En España, aparte de un proyecto de registro internacional (*http://odisea.ciepi.org*) contamos aún con muy poca experiencia en este tipo de repositorios de datos abiertos de carácter científico si bien se pueden mencionar algunos en el campo de las ciencias puras (**Wulff-Barreiro**, 2011; **Arano-Poggi** *et al.*, 2011), y en ciencias sociales y humanidades el trabajo del *Centro de Estudios Avanzados en Ciencias Sociales* (*Ceacs*), del *Instituto Juan March en Estudios e Investigaciones*, compuesta de datos primarios y secundarios de encuestas y estadísticas procedentes del *CIS*, *INE*, la *OCDE* y otros organismos e integrada en la red *Dataverse* de la *Harvard University*. A ello se suma la colección "Recursos i dades primàries", integrada en el repositorio digital de la *Univ. Pompeu Fabra*. A nivel internacional, el repositorio en ciencias sociales más importante es del *Interuniversity Consortium for Political and Social Research* (*Icpsr*) de Estados Unidos.

> Los problemas son la actitud recelosa de los investigadores hacia la compartición de los datos y la falta de formación para su correcta gestión

Plan de gestión de datos científicos

La infraestructura tecnológica de los repositorios de datos no difiere sustancialmente de la existente para otros repositorios. Los grandes problemas para la puesta en marcha, como ya se señaló, son dos: la actitud recelosa de los investigadores hacia la compartición de esos datos y la falta de formación de investigadores y bibliotecarios o gestores de datos para su correcta gestión. **Michener** y **Jones** (2012) resumen en 8 fases el ciclo de vida de los datos de investigación, fases que pueden ser concurrentes o pueden repetirse durante una investigación: planificación, recolección, control de calidad, descripción, preservación, descubrimiento, integración y análisis.

Planificación

Marca cuáles serán los procesos y recursos para completar el ciclo de vida de los datos. Es la fase en la que se marca cuáles son los objetivos del proyecto de conservación y/o explotación de datos, la que define cómo se gestionarán, cuáles serán las políticas y cómo será el plan de sostenibilidad. Tanto la herramienta del *Digital Curation Centre*[8] como la de *DMP tool* de la *University of California*[9] sirven a los investigadores para desarrollar su plan. Probablemente sea la fase más crítica, en donde se debe determinar:

- tipos de datos que serán producidos (brutos, experimentales, imágenes, modelos…);
- cuándo, cómo y dónde se recogerán;
- formatos admitidos y convenciones de nombres;
- si habrá versiones diferentes;
- metadatos obligatorios;
- políticas para compartir;
- cómo deberán ser citados;
- privacidad, seguridad y derechos de autor;
- costes, etc.

Recolección

Recolección efectiva de los datos, de acuerdo con lo planificado, bien de forma automática desde sensores hacia hojas de cálculo o de forma manual mediante la introducción de datos con un teclado o como se haya decidido. La recolección debería ajustarse a los formatos, las unidades de medida, vocabularios, nombres, etc., que se hayan decidido.

Control de calidad

Mecanismos para prevenir errores a la hora de introducir los datos recolectados. Puede consistir en realizar entradas dobles de datos, separadamente por dos personas, o en procedimientos y algoritmos que permiten detectar errores fácilmente.

Descripción de datos

Metadatos que permitan comprender el contenido, formato y contexto de un producto de datos. Normalmente debe describir quién creó, recolectó y gestionó los datos, el contenido y el formato, cuándo y dónde se recolectaron, y –si es relevante- en qué condiciones se almacenaron, cómo se generaron, procesaron, verificaron y analizaron, y por qué se generaron, con qué fin, cuál era el proyecto a partir del cual se generaron.

Preservación

A corto plazo y a largo plazo, con políticas de verificación de la integridad y de copias de seguridad, usando formatos estables de ficheros, asegurando la fiabilidad de los medios de almacenamiento, identificando los datos con valor para el largo plazo, definiendo políticas de depuración, de asignación de identificadores persistentes y de identificación de posibles datos sensibles.

Descubrimiento

Consta de dos partes: una, hacer los datos realmente accesibles, fuera de ordenadores personales, en sitios web accesibles y en formatos reutilizables, abiertos. Y otra, utilizar todas las herramientas posibles para que los datos puedan ser descubiertos (tesauros, ontologías), depositados en repositorios específicos, etc.

Integración

Consiste en el uso de datos de múltiples fuentes, combinados de tal forma que se puedan analizar. Esta fase se refiere al hecho de que los datos recolectados de, por ejemplo, construcción de viviendas, deberían poder combinarse con datos recolectados sobre la población o sobre equipamientos culturales.

Análisis

Es la fase final y consiste, lógicamente, en la generación de interpretaciones y visualizaciones para identificar patrones, verificar o refutar las hipótesis y mostrar los resultados o descubrimientos de forma comprensible.

Conclusiones

Los repositorios de datos abiertos de investigación contribuyen no sólo a la transparencia, pues permiten comprobar si los métodos y resultados de una investigación se han realizado de acuerdo con la cultura científica de cada área, sino que además permiten avanzar a la ciencia puesto que pueden suponer ahorro de tiempo y dinero al reutilizar recursos producidos por otros. Permite además devolver a la sociedad parte de lo que invierte en ciencia mediante la transferencia hacia empresas innovadoras el uso de datos de forma masiva para la puesta en marcha de servicios sobre esos datos.

Los problemas con este tipo de repositorios no son tanto de infraestructura tecnológica como de: concienciación de los investigadores sobre la importancia de compartir datos, algo que las agencias financiadoras empiezan a tratar de poner solución mediante la obligatoriedad de depositar los datos de una investigación; la falta de planificación de cómo gestionar los datos; y, por último, la falta de formación de investigadores y gestores de información y bibliotecarios para gestionar todo el ciclo de vida de los datos científicos.

Notas

1. *http://opendefinition.org*

2. *http://5stardata.info*

3. *http://politica.elpais.com/politica/2011/09/09/actualidad/1315584504_266528.html*

4. *http://ec.europa.eu/research/era/docs/en/others-9.pdf*

5. *http://www.e-irg.eu/images/stories/e-irg_dmtf_report_final.pdf*

6. *http://www.nsf.gov/cise/cise_dmp.jsp*

7. **Greenberg, Jane** (2012). "Bibliotecas digitales para datos de investigación". *Máster en Bibliotecas y Servicios de Información Digital*, Universidad Carlos III, Madrid, 13 febrero. *http://163.117.69.23/mediasite/Viewer/?peid=1c8617e82d83456c940479af3c3f368b*

8. *https://dmponline.dcc.ac.uk*

9. *DMP Tool*, *http:/dmp.cdlib.org*

Bibliografía

Arano-Poggi, Silvia; **Martínez-Ayuso, Gemma**; **Losada-Yáñez, Marina**; **Villegas-Montserrat, Marta**; **Casaldàliga-Riera, Anna**; **Bel-Rafecas, Núria** (2011). "La comunidad 'Recursos y datos primarios' de la *Universitat Pompeu Fabra*: los repositorios institucionales como infraestructuras científicas: estudio de caso". *Revista española de documentación científica*, v. 34, n. 3, pp. 385-407. *http://dx.doi.org/10.3989/redc.2011.3.834*

Berners-Lee, Tim (2009). *Linked data - Design issues*. *http://www.w3.org/DesignIssues/LinkedData.html*

Björk, Bo-Christer; **Welling, Patrik**; **Laakso, Mikael**; **Majlender, Peter**; **Hedlund, Turid**; **Guðnason, Guðni** (2010). "Open access to the scientific journal literature: situation 2009". (Enrico Scalas, Ed.) *PloS one*, v. 5, n. 6, e11273. *http://dx.doi.org/10.1371/journal.pone.0011273*

Fanelli, Daniele (2009). "How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data". (T. Tregenza, Ed.) *PloS one*, v. 4, n. 5, e5738. *http://dx.doi.org/10.1371/journal.pone.0005738*

Gargouri, Yassine; **Lariviere, Vincent**; **Gingras, Yves**; **Brody, Tim**; **Carr, Les**; **Harnad, Stevan** (2012). *Testing the Finch hypothesis on green OA mandate ineffectiveness*. *http://arxiv.org/abs/1210.8174*

Ginsparg, Paul (2011). "ArXiv at 20". *Nature*, n. 476 (7359), pp. 145-147. *http://dx.doi.org/10.1038/476145a*

Krugman, Paul (2013). "La depresión del Excel". *El país*. Ediciones El País, 21 de abril. *http://economia.elpais.com/economia/2013/04/19/actualidad/1366398440_370422.html*

Martin, Ben R. (2013). "Whither research integrity? Plagiarism, self-plagiarism and coercive citation in an age of research assessment". *Research policy*, v. 42, n. 5, pp. 1005-1014. *http://dx.doi.org/10.1016/j.respol.2013.03.011*

Méndez, Eva; **Greenberg, Jane** (2012). "Datos enlazados para vocabularios abiertos y marco general de HIVE". *El profesional de la información*, v. 21, n. 3, pp. 236-244. *http://www.elprofesionaldelainformacion.com/contenidos/2012/mayo/03_esp.pdf*

Open Knowledge Foundation (2012). *The open data handbook*. *http://opendatahandbook.org*

Pampel, Heinz; **Vierkant, Paul**; **Scholze, Frank**; **Bertelmann, Roland**; **Kindling, Maxi**; **Klump, Jens**; **Goebelbecker, Hans-Jürgen** *et al.* (2013). "Making research data repositories visible: the re3data.org registry". *PeerJ preprints*, 1, e21, v. 1. *http://dx.doi.org/10.7287/peerj.preprints.21v1*

Ríos-Hilario, Ana; **Martín-Campo, Diego**; **Ferreras-Fernández, Tránsito** (2012). "*Linked data* y *linked open data*: su implantación en una biblioteca digital. El caso de *Europeana*". *El profesional de la información*, v. 21, n. 3, pp. 292-297. *http://dx.doi.org/10.3145/epi.2012.may.10*

Schumpeter (2011). "Building with big data". *The economist*. *http://www.economist.com/node/18741392*

Torres-Salinas, Daniel (2009). "Compartir datos (data sharing) en ciencia: el contexto de una oportunidad". *Anuario ThinkEPI*, v. 4, pp. 262-265. *http://www.thinkepi.net/compartir-datos-data-sharing-en-ciencia-el-contexto-de-una-oportunidad*

Torres-Salinas, Daniel; **Robinson-García, Nicolás**; **Cabezas-Clavijo, Álvaro** (2012). "Compartir los datos de investigación en ciencia: introducción al data sharing". *El profesional de la información*, v. 21, n. 2, pp. 173-184. *http://dx.doi.org/10.3145/epi.2012.mar.08*

Wulff-Barreiro, Enrique (2011). "Approaches to open data for science in Spain". *Data science journal*, n. 10, pp. 13-23. *http://www.jstage.jst.go.jp/article/dsj/10/0/10_13/_article*