

ORDER

HERE



ANA B. RÍOS HILARIO, YOLANDA MARTÍN GONZÁLEZ

La publicación de los datos gubernamentales según el “LOD cloud diagram”

Asistimos, en la actualidad, al nacimiento de un nuevo modelo de gobierno, denominado *open government* o *gobierno abierto*. Se trata de una filosofía convertida en doctrina política y en eje director de muchas de las iniciativas puestas en marcha por las administraciones públicas en el ámbito nacional e internacional. Esta nueva forma de administración se encuentra estrechamente relacionada con otros conceptos y movimientos que confluyen en la actualidad, como son la administración electrónica, la rendición de cuentas en la gestión pública, la participación y la colaboración ciudadana, la transparencia y el acceso a la información.

Una de las consecuencias más notables del gobierno abierto es la apertura de los datos que obran en poder del sector público que supone en realidad, una nueva forma de responder a las demandas democráticas de los ciudadanos al tiempo que un reconocimiento del potencial principalmente económico y de empleabilidad que conlleva la reutilización de la información liberada por los organismos gubernamentales.

El acceso a los datos públicos abiertos –*open government data*– se facilita mediante unos instrumentos creados expresamente para ello y regulados en las

Ana B. Ríos Hilario (email: anarihi@usal.es)
 Yolanda Martín González (email: ymargon@usal.es)
 Dpto. Biblioteconomía y Documentación, Universidad de Salamanca

Recibido: 31-08-2015. Aceptado: 10-09-2015

Citación: Ríos Hilario, Ana B.; Martín González, Yolanda (2015). “La publicación de los datos gubernamentales según el ‘LOD cloud diagram’”. *Tabula*, n. 18, pp. 85-102

correspondientes normativas que son los denominados catálogos o repositorios de información pública, puestos en marcha por los gobiernos a nivel internacional, nacional, regional y/o local. No obstante, en los últimos años se insiste en la necesidad de dar un paso más y enlazar dichos datos de modo que puedan compartirse, para lo cual se requiere que se alojen en un repositorio universal como es la Nube.

Y es que de la sociedad de la información hemos pasado a la sociedad del dato, de la web de los documentos a la web de los datasets en la que los diferentes datos se conectan y vinculan entre sí a través de una serie de pautas perfectamente establecidas que dan lugar a la tecnología denominada LOD (*linked open data*, datos abiertos y vinculados). Los datos así publicados se recogen en la denominada "nube de datos".

Tomando como base las anteriores premisas, se estimó oportuno la realización de un análisis que expusiera la disposición actual y representación de los *datasets* perteneciente a la categoría denominada "Gobierno" presentes en la *LOD cloud diagram* (nube de los datos vinculados). Este estudio se efectúa desde la perspectiva de la gestión y publicación de los datos, que desde aquí abogamos que dicha tarea se ejecute por el gestor de la información (en este caso, *data manager*) de estas instituciones públicas, ya que son estos profesionales los especialistas en la provisión y asignación de vocabularios y metadatos.

Partiendo de este objetivo principal, se establecieron los siguientes propósitos específicos:

- Explicar los conceptos relacionados con el movimiento *open government data*, *linked data* y *linked open data*.
- Describir los requisitos que conlleva la liberación y vinculación de los datos del sector público.
- Explicar el LOD cloud diagram centrándonos en su definición, objetivos y formación.
- Analizar de modo específico las características generales de los conjuntos de datos procedentes de la administración pública, en función de las siguientes variables: conjuntos de datos disponibles y cumplimientos de una serie de recomendaciones en lo referente a la interconexión, utilización de vocabularios y asignación de metadatos.

En función de los objetivos anteriormente propuestos se aplicó para la realización de la investigación una doble metodología. Para la elaboración del estado de la cuestión de los conceptos a estudiar se optó por la consulta de diversas fuentes documentales; mientras que para el análisis específico de los datasets se ha aplicado una metodología fundamentalmente de carácter descriptivo a través del examen detallado de la propia nube de datos vinculados.

A partir de los objetivos planteados y de la metodología establecida, el texto se estructura en las siguientes partes. Tras la introducción, los primeros apartados están destinados a la definición de los conceptos *open government data* y *linked open data*. A continuación figura el análisis propiamente dicho de los conjuntos de datos públicos, presentes actualmente en la nube, en función de las variables anteriormente indicadas. Finaliza el artículo resaltando las principales conclusiones obtenidas y haciendo referencia a la documentación empleada para la realización del mismo.

Del Open Government Data (OGD) al Linked Open Data

Es muy frecuente hoy en día la utilización de la expresión “data” en las diversas formas de comunicación científica, generalmente, acompañada de otra palabra que precise su significado, como puede ser *open data*, *data mining* y el más reciente *big data*. Por su parte, el adjetivo “open” en conjunción con otros vocablos ha dado lugar a toda una serie de iniciativas como *open access*, *open government* u *open science*, agrupadas bajo la denominación de “movimientos abiertos”.

Según el *Open Definition Advisory Council* (2014) el concepto *open* haría referencia tanto al “contenido incluido en música, películas y libros; los datos de carácter científico, histórico, geográfico; o cualquier otro tipo de información gubernamental y de otras administraciones públicas”. Como podemos comprobar, la materia objeto de estudio en este texto se incluye dentro de la propia definición oficial, que se realiza sobre la expresión anteriormente indicada. Esta misma institución expresa las condiciones de distribución que debe cumplir una obra para que sea considerada abierta. De modo específico enumera las siguientes 11 pautas: acceso, 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.

La primera vez que se utiliza este calificativo es dentro de la expresión *open source* que comienza a utilizarse a finales de la década de los 90 y, desde entonces, cada vez es más extenso el elenco de enunciados que incluyen el término *open*. El código abierto no sólo significa el acceso al código fuente (Open Source Initiative, 1998) sino que deben cumplir toda una serie de criterios especificados en la propia iniciativa.

Open Government Data

Uno de los movimientos que surgen entorno al concepto de abierto es, como ya hemos señalado, el *open government data*. Este aparece motivado por el importante desarrollo digital sufrido por la Administración y por el cambio en las formas de

gestión en el sector público que se materializan hoy día en formas de actuar mucho más modernas y efectivas. Por otra parte, surge como consecuencia de un nuevo modelo de gobierno, *open government* o *gobierno abierto*, basado en la idea de que la Administración y los responsables políticos han de rendir cuentas ante el ciudadano y éste a su vez ha de participar y colaborar en los asuntos públicos.

Aunque se desconoce el origen concreto de esta nueva filosofía política, éste suele situarse en Estados Unidos donde, el 21 de enero de 2009, el Presidente Obama dio un fuerte impulso a su iniciativa al hacer público su Memorando sobre Transparencia y Gobierno Abierto. Otros países como Australia, Nueva Zelanda o el Reino Unido también han sido pioneros en la apuesta por esta nueva forma de gestionar los asuntos públicos.

El modelo de gobierno abierto se sustenta sobre los principios de transparencia, colaboración y participación ciudadana. En concreto, el principio de transparencia propugna, entre otras cosas, la liberación de la información y los datos que obran en poder de la Administración y que se disponen en formato libre de modo que puedan ser utilizados, reutilizados y distribuidos por terceros. Asimismo este principio persigue la elaboración de marcos normativos que regulen el derecho de acceso a la información en aquellos ámbitos donde aún no lo está. De este modo, las políticas en materia de gobierno abierto se materializan en iniciativas o estrategias, entre las que destaca, la apertura de datos públicos abiertos –*open government data*– y su consiguiente reutilización.

Y es que el sector público, en el cumplimiento de sus funciones, genera, gestiona y conserva una gran cantidad de información de calidad. Dicha información presenta un importante potencial socioeconómico y su difusión contribuye a la transparencia en la gestión pública y a la generación de confianza por parte de la ciudadanía. Tal es así, que los ciudadanos se convierten en demandantes de información y la Administración en una de las mejores proveedoras de este activo abundante y barato (Mendo y otros, 2013).

La finalidad de las iniciativas *open government data* consiste en hacer públicos los datos en poder de las administraciones para que puedan ser posteriormente utilizados, distribuidos y reutilizados con la finalidad de generar nuevos servicios, productos y aplicaciones que, en muchos casos, añadirán un valor al dato primario y contribuirán a potenciar la innovación en cualquier país.

Es tal la relevancia que está adquiriendo esta filosofía política que los líderes de los países que conforman el G-8, en la reunión que celebraron en junio de 2013, aprobaron la *Open Data Charter*. En este documento se contempla un conjunto de principios básicos relativos a la liberación y reutilización de datos facilitados por los gobiernos del G8 y que se resumen en las siguientes ideas: los datos son públicos por definición; debe existir un equilibrio entre calidad y cantidad de los datos ofrecidos; los datos podrán ser utilizables por todos; y la apertura de los datos deberá mejorar la gobernanza y favorecer la innovación.

Para la Unión Europea, la apertura de los datos y su consiguiente reutilización se entienden como una de las políticas de fomento de la Sociedad de la Información en la línea de la denominada *Estrategia de Lisboa*, en la que se insistía en aprovechar las nuevas tecnologías y la nueva realidad de la sociedad del conocimiento para fomentar el empleo y el crecimiento económico. Es decir, que los países adheridos a esta iniciativa se plantean desde el inicio el acceso a sus datos, no sólo como una forma de hacer más transparentes y abiertas las administraciones públicas, sino también y fundamentalmente, como una forma de conseguir ayudar a las empresas a aprovechar el potencial de esta información, y contribuir así al desarrollo económico y al fomento del empleo. La propia Comisión Europea afirmó en 2011 que “las administraciones públicas europeas tienen una auténtica mina de oro de potencial económico sin aprovechar” (Comisión Europea, 2011).

El movimiento *open data* o de datos abiertos se puede explicar cómo una filosofía y, al mismo tiempo, una práctica que persigue que determinados datos estén disponibles de forma libre a todo el mundo, en formatos que permitan que puedan ser utilizados, reutilizados y redistribuidos y sin ningún tipo de restricción legal, financiera o tecnológica y al mismo tiempo, como una forma de comunicación abierta, permanente y bidireccional entre la administración y los ciudadanos (Ferrer-Sapena; Peset; Aleixandre-Benavent, 2011).

No obstante, no toda la información generada por el sector público posee las mismas características y despierta el mismo interés de cara a su apertura y reutilización. La Organización para la Cooperación y el desarrollo Económico (OCDE), en su informe *Participative Web: User-Created content* (DSTI/ICCP/IE 2) elaborado en 2005, realiza una distinción entre dos grupos de información pública (cit. por Ramos y Arquero, 2012). Por un lado, la información del sector de las administraciones producida y recopilada por los organismos del sector público como parte de su actividad, se considera información dinámica, continua y fácilmente accesible. Esta información, directamente generada por el sector público y asociada al funcionamiento de su actividad (p.e. datos económicos, geográficos, estadísticos, jurídicos, etc.) se halla vinculada al concepto de conocimiento como recurso competitivo en la sociedad de la información. Por otra parte, existen otros contenidos públicos más estáticos, conservados, registrados y guardados –más que producidos– en los organismos públicos que no están asociados directamente a la función de gobierno, tampoco tienen un específico valor comercial, aunque se orientan a otros propósitos públicos, por ejemplo, culturales o educativos, como es el caso de las instituciones GLAM (*galleries, libraries, archives and museums*).

Este proceso de apertura de los datos públicos debe realizarse bajo una serie de principios establecidos en 2008, por 30 defensores del gobierno abierto, y que son las siguientes (Sunlight Foundation, 2010):

1. La información tiene que ser completa.
2. Los datos deben provenir directamente de la fuente, sin modificación ni agregación alguna.
3. La publicación debe realizarse a tiempo para preservar el valor de la información.
4. La información tiene que estar disponible para el mayor número de usuarios y propósitos posibles.
5. Los datos deben presentar una estructura mínima que permita su tratamiento automatizado.
6. La publicación tiene que estar libre de discriminación y disponible para cualquier necesidad de registro previo.
7. Los formatos de publicación deben ser no propietarios puesto que ninguna entidad tiene control exclusivo sobre ellos.
8. La información debe ser de uso libre, es decir, no estar sujeta a derechos de autor, patentes, marcas ni a regulaciones de secreto industrial o comercial. Tampoco ha de estar sujeta a restricciones de privacidad, seguridad o privilegio reguladas por las normas.

A estos principios, la Sunlight Foundation añadió en 2010 dos premisas más: los datos deben permanecer en el tiempo y los datos no deben tener costo de uso.

Sin embargo, no todos los datos se pueden liberar y los que se liberan no pueden hacerlo de cualquier forma. Los requisitos que deben cumplir los datos para que pueda ser liberados son las siguientes (Ximdex, 2014):

- Deben ser procesables automáticamente por máquinas (machine readable).
- Deben estar publicados en formatos abiertos que, a poder ser, no dependan de empresas concretas para favorecer así su reutilización (así, por ejemplo, es preferible publicar en formato CSV –*comma separated value*–, en lugar de formato Excel de Microsoft).
- Ha de facilitarse el acceso masivo a los mismos (normalmente “en bruto”).
- Deben tener una “granularidad” suficiente para que resulten útiles al usuario.
- Deben ser fáciles de encontrar en un único punto de acceso en línea, y ser puntualmente publicados (compromiso con los usuarios potenciales) y en estándares abiertos.
- Deben poder ser descargables desde diferentes tipos de terminales y navegadores.

En 2010, Berners Lee desarrolló un sistema de clasificación de 5 estrellas “con el fin de animar a la agentes implicados—especialmente a los propietarios de los datos del gobierno— en el camino hacia los buenos datos vinculados” (figura 1).






★	Datos disponibles en la web, en cualquier formato, pero con una licencia abierta para ser <i>open data</i> (por ejemplo, un pdf)	
★★	Datos en formato estructurado (por ejemplo un Excel)	
★★★	Datos en un formato no propietario: por ejemplo, ficheros de datos separados por comas (<i>comma separated values, CSV</i>)	
★★★★	Uso de un estándar W3C (RDF y SPARQL) para identificar las cosas, de modo que otras personas puedan apuntar a esas cosas	
★★★★★	Enlazar unos datos con otras fuentes de datos para dotarlos de contexto	

Figura 1. Clasificación “5 estrellas”

Fuente: Elaboración propia a partir de la información de Tim Bernes Lee, 2010

En el ámbito del *open government data* hay un aspecto muy importante que son las licencias que establecen las condiciones de uso de los datos. Las licencias de datos libres otorgan, básicamente, las siguientes libertades (Open Linked Data, guía para la liberación de datos, 2013):

- a. Plena libertad para compartir (copiar, distribuir) y utilizar los datos,
- b. Plena libertad para crear obras derivadas de esos datos, y
- c. Plena libertad para adaptarlos (modificarlos y transformarlos).

La diferencia entre licencias proviene de las "obligaciones" que imponen al usuario de los datos, distinguiéndose dos tipos principales:

- a. "Dominio Público", sin exigir ninguna contraprestación a cambio.
- b. "Open Data License", exigiendo al usuario de los datos que:
 - Se realice una correcta atribución del origen de los datos,
 - Se ofrezcan los datos derivados con el mismo tipo de licencia,
 - Se ofrezca una versión "libre" paralela a cualquier versión restringida.

Además, la apertura de datos públicos requiere de unos estándares para que un concepto en una administración y en un país signifique lo mismo en otro y que los esquemas en los que está codificada la información, la base de datos, sean interoperables. La interoperabilidad permite que la información fluya sin ningún tipo de problema –fundamentalmente tecnológico– por distintos sistemas, haciendo posible que diferentes organizaciones puedan trabajar conjuntamente. La prioridad debería ser, en primer lugar, que los datos disponibles estén "en bruto" (*raw data*), procedentes de tablas, bases de datos, etc., y en un formato informático estándar y abierto que permita que sean interpretados fácilmente para mejorarlos más adelante con tecnologías *linked data*.

Las iniciativas *linked data* y *linked open data*

Si queremos conocer de modo preciso el significado del término *linked data* debemos acudir a la página oficial, denominada del mismo modo, en la que se define este concepto como "la utilización de la Web para conectar los datos relacionados que no estaban vinculados previamente, o el uso de la Web para disminuir los obstáculos en la conexión de los datos actualmente vinculados mediante otros métodos" (*Linked data*, 2015). Dicho de otro modo, LD sería "la publicación de datos estructurados que permiten la conexión y enriquecimiento de los metadatos" (Ríos, 2014), de tal forma que "diferentes representaciones de un mismo contenido puedan encontrarse y enlazarse entre recursos relacionados" (*European Linked Open Data*, 2015).

Para hablar de LD los datos deben publicarse de acuerdo con los principios diseñados para facilitar los vínculos entre los conjuntos de datos, elementos y vocabularios controlados (Berners Lee, 2006). Estas prácticas fueron presentadas en 2006 por Tim Berners Lee y se han dado a conocer como los principios de *linked data*. Tales principios son los siguientes:

1. Usar URIs (*uniform resource identifiers*) para identificar los recursos de forma unívoca.
2. Usar URIs http para poder acceder a la información del recurso.
3. Ofrecer información sobre los recursos usando RDF.

4. Incluir enlaces a otros URIs, facilitando el vínculo entre los distintos datos distribuidos en la web.

Es decir, este conjunto de buenas prácticas o recomendaciones que se incluyen bajo este concepto van a permitir, exponer, compartir y conectar conjuntos de datos, información y conocimiento en la Web Semántica mediante la utilización de URIs para la identificación de las cosas y de estándares como RDF para la descripción de las mismas.

Un paso más en esta iniciativa sería *linked open data* (LOD, datos abiertos y vinculados). Podemos definir LOD como la publicación de datos estructurados, lo que permite que los metadatos se conecten y enriquezcan de tal modo que diferentes representaciones de un mismo contenido puedan encontrarse y enlazarse entre recursos relacionados (*Europeana Linked Open Data*, 2015).

Según Mitchell (2013, p. 12) “*linked open data* comprende dos conceptos distintos, el primero es que los datos publicados en la Web deberían conectarse fácilmente con información (“enlazados”) y al hacer esto debería ser accesible tanto para las computadoras como para los humanos (“datos”). El segundo concepto clave para LOD es que para que los datos se vinculen y se reutilicen, deben estar abiertos y libres de las restricciones de derechos y copyright (“abiertos”)”.

A menudo los términos *linked data* y *linked open data* se utilizan indistintamente, pero existe una matización importante que los diferencia. La etiqueta *linked open data* se utiliza ampliamente pero a menudo para referirse a los datos vinculados en general. Ni todos los datos vinculados estarán abiertos, y ni todos los datos abiertos serán vinculados. Los datos enlazados además serán abiertos si se liberan bajo acuerdos de licencia que permitan su libre reutilización. Por lo tanto, se debe tener cuidado al utilizar la expresión apropiada, dependiendo de los términos de licencia de los datos en cuestión (*Linked data*, 2015).

Para finalizar este apartado teórico nos gustaría reflexionar brevemente sobre los conceptos *open data* y *linked open data* anteriormente definidos. Estas expresiones aunque estén íntimamente relacionadas no son sinónimas: mientras *open data* hace referencia a la interoperabilidad legal de los datos *linked open data* alude a la interoperabilidad técnica de los mismos.

Los datos gubernamentales en el LOD cloud diagram

Definición del LOD cloud diagram

El *LOD cloud diagram* (diagrama de la nube LOD) recoge los *datasets* publicados como datos vinculados que han sido recopilados por los colaboradores del proyecto Linking Open Data en conjunción con otra serie de personas e instituciones (figura 2). El propósito principal de dicho proyecto (Linking Open Data, 2015)

es "extender la Web como un bien común de datos mediante la publicación de varios conjuntos de datos abiertos como RDF en la Web y mediante el establecimiento de enlaces RDF entre elementos de datos de diferentes fuentes". Dicho diagrama se elabora a través de los metadatos recogidos y comisariados por los colaboradores del Data Hub, así como los metadatos extraídos del último rastreo de la web *linked data* realizado en abril de 2014. El Data Hub es "un registro de datos en el que se puede compartir información sobre paquetes de datos de cualquier tipo y describirlos de forma colaborativa" (Isaac; Waites; Young; Zeng, 2011). La elección de esta fuente para la realización de este estudio se debe a su representatividad y actualidad, aunque hay que constatar que no todos los conjuntos de datos publicados como LOD se encuentran en la misma pero si los más relevantes, y que día a día aumenta la publicación de datasets que aplican esta tecnología.

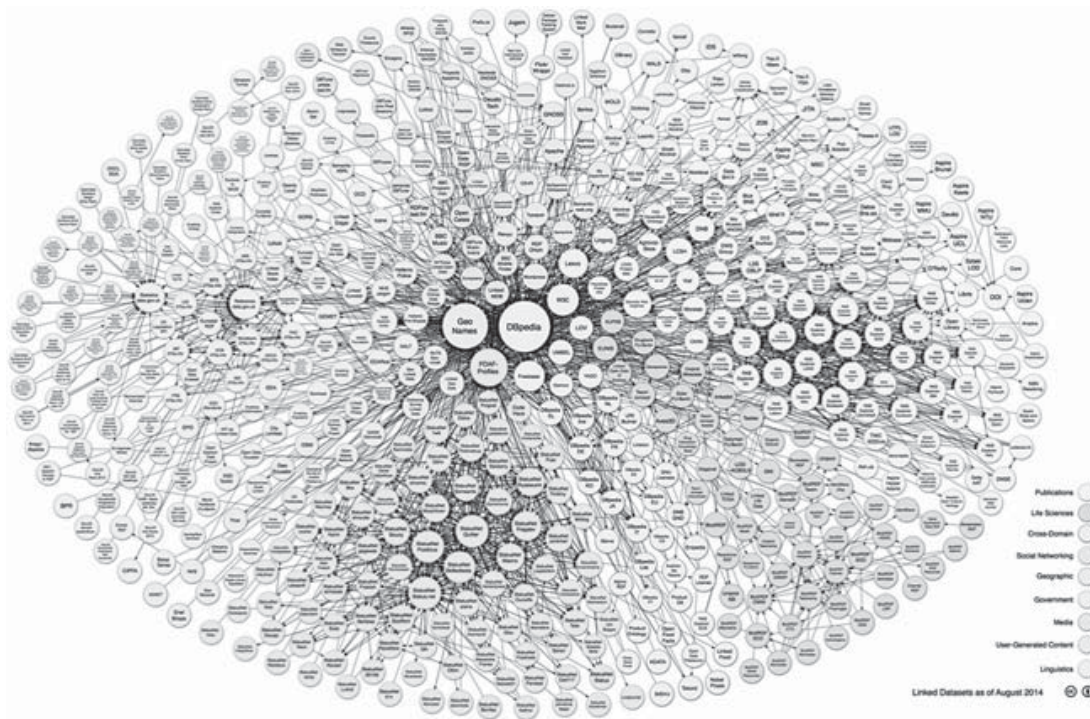


Figura 2. Diagrama de la nube LOD
Fuente: <http://lod-cloud.net/>

- En la nube LOD figuran dos documentos independientes que nos muestran:
- El estado de la nube en 2014: documento que presenta información sobre la estructura y el contenido del subconjunto rastreado de la nube LOD en abril de 2014.

Descripción de los datasets gubernamentales en el LOD cloud diagram

Linked data se está aplicando para compartir datos referentes a un amplio abanico de dominios de diferentes campos. La tabla que figura a continuación presenta una visión general de los 1.014 conjuntos de datos que se identificaron en el último rastreo clasificados por categorías temáticas.

Tema	Conjunto de datos	%
Gobierno	183	18,05%
Publicaciones	96	9,47%
Ciencias de la vida	83	8,19%
Contenido generado por el usuario	48	4,73%
Dominios cruzados	41	4,04%
Medios	22	2,17%
Geografía	21	2,07%
Web social	520	51,28%
TOTAL	1.014	

Tabla 1. Conjuntos de datos por categorías temáticas
Fuente: <http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/>

La categoría denominada "Gobierno" comprende los conjuntos de datos publicados como "datos vinculados por parte de los gobiernos federales o locales, incluyendo una gran cantidad de conjuntos de datos estadísticos". Ejemplos destacados dentro de esta categoría serían los datasets del gobierno británico (data.gov.uk) y la comunidad de conjuntos de datos provenientes de los gobiernos locales del Reino Unido (opendatacommunities.org) (Schmachtenberg; Bizer; Paulheim; 2014). En el caso español destacan los conjuntos relativos a los gobiernos autonómicos de Asturias y País Vasco, y los datos locales de Zaragoza.

Como puede observarse en la tabla anterior de las 8 materias identificadas, la categoría de los datos de las administraciones públicas ocupa la segunda posición, con 183 conjuntos de datos identificados, lo que se corresponde con el 18,05% del espacio en la nube. Destaca muy por encima del resto los conjuntos de datos relativos a la web social (520) que ostenta la primera posición.

Para un análisis más exhaustivo de la realización del estado de la nube ejecutado en abril de 2014 podemos consultar el texto titulado *Adoption of the linked*

data best practices in different topical domains (Schmachtenberg; Bizer; Paulheim, 2014) o el resumen del mismo que se establece en el documento *State of the LOD cloud 2014*.

Es importante destacar el apartado 4 del primer documento destinado a una serie de buenas prácticas o recomendaciones y que, a continuación, vamos a analizar detalladamente, ya que consideramos que no sólo es importante ver el estado actual de los conjuntos de datos vinculados de los gobiernos y su posición con respecto a las otras áreas, sino que también esta radiografía podrá ser interesante y útil para aquellas instituciones públicas que estén interesadas en publicar sus datos como LOD y para los responsables de ejecutar esta tarea.

LD se fundamenta en que “los editores de datos soporten aplicaciones para el descubrimiento e integración de los datos a través del cumplimiento de un conjunto de buenas prácticas en las áreas de la vinculación, uso de vocabularios y provisión de metadatos” (Schmachtenberg; Bizer; Paulheim, 2014). En la citada obra se expone en qué medida los datos analizados implementan estas buenas prácticas.

El primer apartado hace referencia a la interconexión de los datos. A través de los enlaces RDF, los proveedores de datos conectan sus conjuntos de datos en un solo gráfico global que se puede navegar mediante las aplicaciones correspondientes y permitir el descubrimiento de datos adicionales siguiendo los enlaces RDF. Del total de conjuntos rastreados el 56,11% enlaza al menos a otro conjunto de datos, el resto de conjuntos son sólo objetivos de enlaces RDF. Dentro de la información proporcionada en este punto, es interesante resaltar los predicados más utilizados para la interconexión en el caso de los datos de las administraciones. Como podemos comprobar en la tabla 2 son los términos “editor” y “cobertura espacial” del esquema Dublin Core y la propiedad “sameAs” (igual como) del Web Ontology Language (OWL) los predicados más empleados dentro de este ámbito.

Categoría	Predicado	Uso
Gobierno	dct:publisher	47,57%
	dct:spatial	30,10%
	owl:sameAs	24,27%

Tabla 2. Predicados más utilizados para la interconexión

En lo referente a los vocabularios empleados se especifica que en la medida de lo posible se utilicen para representar los datos términos de vocabularios ampliamente desarrollados (Schmachtenberg; Bizer; Paulheim, 2014). Se establece

una diferenciación entre los vocabularios propietarios y los vocabularios de-referenciales. Un vocabulario es propietario si se utiliza exclusivamente por un conjunto de datos, mientras que, el concepto de vocabulario no propietario se aplica cuando como mínimo dos conjuntos de datos emplean el mismo vocabulario.

Los tres vocabularios más utilizados en la categoría de datos gubernamentales son: *Dublin Core Metadata Initiative Terms* (dcterm), *Data Cube Vocabulary* (cube) y *Open Data Communities* (odc) (tabla 3).

Categoría	Vocabulario	Uso
Gobierno	dct	63,98%
	cube	60,75%
	odc	46,24%

Tabla 3. Utilización de vocabularios

Para los vocabularios propietarios es esencial que sean de-referenciales y enlazables con otros vocabularios, de este modo los agentes pueden interpretar sus semánticas. La siguiente tabla nos presenta los siguientes parámetros: vocabularios propietarios, conjuntos de datos con vocabularios propietarios y nivel de de-referencialidad que puede ser completa, parcial o nula.

Categoría	VP usados	Conjuntos de datos VP	De-referenciabilidad		
Gobierno	48 (12,70%)	35 (18,82%)	Completa 20,83%	Parcial 12,50%	Ninguna 66,67%

Tabla 4. Uso y de-referenciabilidad de los vocabularios empleados

El cuarto y último punto se centra en el estudio de los metadatos empleados en función de las siguientes variables: suministro de información de procedencia, información de licencias, provisión de metadatos a nivel de dataset o conjuntos de datos y provisión metadatos de acceso alternativo.

En lo referente a la información de procedencia los datos totales indican que el 41,40% de todos los conjuntos de datos de esta categoría utilizan un vocabulario de procedencia. En cuanto a los vocabularios individuales el 100% utiliza DC o DCTerm no dándose ningún caso del empleo de MetaVocab (vocabulario

RDF para la descripción de autores). En un ínfimo porcentaje (1,30%) se emplea *prv* o *prov* (*Provenance Vocabulary Core Ontology*) (tabla 5).

Categoría	Cualquier vocabulario de procedencia	Uso DC	Uso MV	Uso prv o prov
Gobierno	77 (41,40%)	100%	0,00%	1,30%

Tabla 5. Información de procedencia

El siguiente ítem especifica el suministro de información sobre licencias de uso de datos. En total, tan sólo el 7,85% de todos los conjuntos de datos proporcionan información de licencias en RDF. El caso particular de los datos gubernamentales es la categoría que más información proporciona sobre las licencias empleadas con un 29,47%. A continuación se sitúan las categorías de dominios cruzados y de contenido generado por el usuario con un 11,36% y 10,91% respectivamente.

El tercer subapartado especifica la provisión de metadatos a nivel de *dataset* mediante la utilización del vocabulario VoID (*Vocabulary of Interlinked Datasets*), ya sea como declaraciones en línea en el conjunto de datos o en un archivo VoID separado. Si examinamos los datos de la categoría objeto de estudio (tabla 6) vemos como de la totalidad de *datasets* 75 conjuntos de datos utilizan el vocabulario VoID de los que el 6,99% presentan un mecanismo de retroenlace y el 3,23% facilita una dirección bien reconocida. El mayor porcentaje (31,18%) presenta el empleo de VoID en línea.

Categoría	Total	Enlace	Bien reconocido	En línea
Gobierno	75 (40,32%)	6,99%	3,23%	31,18%

Tabla 6. Metadatos a nivel de “dataset”

Por último, de la totalidad de conjunto de datos presentes en la categoría Gobierno, 61 (32,80%) emplean métodos de acceso alternativos. Por su parte, el 30,31% muestra *endpoints* SPARQL y con un porcentaje prácticamente igual (30,65%) figura Dump. En este sentido, diremos que SPARQL (*Protocol and RDF Query Language*) es un lenguaje estandarizado para la consulta de grafos RDF, mientras que Dump hace referencia al volcado de los datos (tabla 7).

Categoría	Métodos alternativos	SPARQL	Dump
Gobierno	61 (32,80%)	30,11%	30,65%

Tabla 7. Métodos de acceso alternativos

Conclusiones

El movimiento *open government data* responde a una nueva forma de gobierno en la que se aboga por el acceso a la información como medida de transparencia en la gestión pública pero también de fomento de la innovación, la creación de empleo y de impulso económico.

Tal como se ha explicado en líneas anteriores, para que el *open government data* se lleve a cabo, los datos en poder de los organismos públicos deben liberarse cumpliendo una serie de requisitos establecidos a nivel internacional, incluyendo licencias que indiquen las condiciones de uso de los mismos y respetando unos estándares que permitan la codificación de la información a través de esquemas comunes, de modo que puedan enlazarse unos *datasets* con otros.

En el presente trabajo se ha intentado obtener una "fotografía" del estado de los *datasets* pertenecientes a la categoría denominada "Gobierno" de la nube de datos vinculados. Tal análisis nos ha permitido extraer un conjunto de buenas prácticas que pueden servir de ayuda a los gestores de datos para la representación de este tipo de *datasets*.

Tras el análisis realizado, en primer lugar, habría que destacar el lugar predominante que ostenta esta categoría dentro del total de conjuntos de datos, lo que pone de relieve la importancia que para los distintos organismos públicos tiene la aplicación de las tecnologías LOD. Por otro lado, como predicado más utilizado para la interconexión de los conjuntos figura "dct:Publisher". En lo referente al empleo de vocabularios propietarios subrayar el uso de DCMI Terms, si bien es cierto, que lo más destacable dentro de este apartado es la aplicación de un vocabulario exclusivo de este grupo –*Open Data Communities*– utilizado por los gobiernos locales que pertenecen a la comunidad que recibe la misma denominación. Asimismo, destaca el predominio absoluto de Dublin Core como vocabulario de referencia. Como nota muy positiva para los datos públicos, es que se sitúan en el primer puesto en lo referente a la información que proporcionan acerca de las licencias empleadas. A nivel de *dataset* prevalece el uso del vocabulario VoID en línea. Finalmente, entre los métodos de acceso alternativos se obtienen porcentajes similares entre el empleo de SPARQL y Dump.

Nos gustaría concluir este apartado proponiendo posibles investigaciones en relación con la temática expuesta en la presente comunicación. En primer lugar, sería conveniente realizar un estudio exhaustivo de cada uno de los *datasets* gubernamentales presentes en el *Mannheim Linked data Catalog* (2014). Dicho instrumento proporciona información detallada sobre los conjuntos de datos disponibles en la Web, en el momento de la realización del último estudio del LOD cloud diagram. De igual modo, sería interesante realizar un análisis comparativo de los conjuntos de datos públicos en relación con el resto de las categorías de datos incluidas en la nube de los datos vinculados. En concreto, podría estudiarse su relación con la categoría denominada Publicaciones en la que se incluyen los datos pertenecientes a las instituciones GLAM (*galleries, libraries, archives and museum*).

Para finalizar, quisiéramos señalar la necesidad de que la función de publicación y descripción de los *datasets* gubernamentales sea realizada por el especialista de la información de dichas instituciones, es decir, que se lleve a cabo por el *data manager* de la organización y que dicha publicación se realice siguiendo el conjunto de buenas prácticas anteriormente expuestas.

Bibliografía

- COMISIÓN EUROPEA (2011). “Agenda Digital: hacer fructificar los datos de las administraciones”. <http://europa.eu/rapid/press-release_IP-11-1524_es.htm> [Consulta: 04/06/2015]
- BERNES-LEE, T. (2006). “Linked data – Design Issues”. <<http://www.w3.org/DesignIssues/LinkedData.html>> [Consulta: 04/06/2015]
- Europeana Linked Open Data (2015). <<http://data.europeana.eu>> [Consulta: 04/06/2015]
- FERRER SAPENA, A.; PESET, F.; ALEIXANDRE-BENAVET, R. (2011). “Acceso a los datos públicos y su reutilización: Open data y Open Government”. *El profesional de la información*. Vol. 20, n. 3, p. 260-269. <<http://www.elprofesionaldelainformacion.com/contenidos/2011/mayo/03.pdf>> [Consulta: 04/06/2015]
- ISAAC, A.; WAITES, W.; YOUNG, J.; ZENG, M. (2011). “Library Linked data Incubator Group: Datasets, Value Vocabularies, and Metadata Element.” <<http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025>> [Consulta: 04/06/2015]
- Linked Data: connect distributed data across the Web (2015). <<http://linkeddata.org/home>> [Consulta: 04/06/2015]
- LinkingOpenData (2014). <http://www.w3.org/wiki/SweoIG/TaskForces/CommunityProjects/LinkingOpenData> [Consulta: 04/06/2015]
- Open definition (2015). <<http://opendefinition.org/>> [Consulta: 04/06/2015]
- Open Source Initiative (1998). <<http://opensource.org/osd>> [Consulta: 04/06/2015]
- Mannheim Linked data Catalog (2014). <<http://linkeddatacatalog.dws.informatik.uni-mannheim.de>> [Consulta: 04/06/2015]
- MENDO, C. et al. (2013). “Del acceso a la reutilización, del dato al documento: una visión conceptual de la información pública”. *Revista Española de Documentación*

- Científica*. Vol. 36, n.3, p. 1-12. file:///C:/Users/Yolanda/Downloads/805-1695-1-PB.pdf [Consulta: 04/06/2015]
- MITCHEL, E. (2013). "Library Linked data: research and adoption". *Library Technology Reports*. Vol. 49, n. 5, p. 11-25.
- RAMOS SIMÓN, F.; ARQUERO AVILÉS, R. (2012). "Las bibliotecas, archivos y museos ante la nueva directiva de reutilización de información del sector público". En: Congreso Nacional de Bibliotecas Públicas (6°. Burgos. 2012). *VI Congreso Nacional de Bibliotecas Públicas/ Europeana Conference*. Retrieved <<http://www.mcu.es/bibliotecas/MC/2012/CongresoBP/comunicaciones.html>> [Consulta: 04/06/2015]
- RÍOS HILARIO, A. B. "Linked Open Data" (2014). En: *DINLE: Diccionario Digital de Nuevas Formas de Lectura y Escritura*. Salamanca: Ediciones Universidad de Salamanca; [S.l.]: Red Internacional de Universidades Lectoras. <<http://dinle.eusal.es/searchword.php?valor=Linked+Open+Data>> Consulta: [04/06/2015]
- SCHMACHTENBERG, M.; BIZER, C.; PAULHEIM, H. (2014) . "Adoption of the linked data best practices in different topical domains". <<http://dws.informatik.uni-mannheim.de/fileadmin/lehrstuehle/ki/pub/SchmachtenbergBizerPaulheim-AdoptionOfLinkedDataBestPractices.pdf>> Consulta: [04/06/2015]
- SUNLIGHT FOUNDATION (2010). "Ten Principles for opening up government information". <<http://sunlightfoundation.com/policy/documents/ten-open-data-principles>> Consulta: [04/06/2015]
- XIMDEX (2014). "Linked Open Data. Guía para la publicación de datos abiertos y enlazados". < http://www.ximdex.com/Guides/OXE_OpenData_Guide_ES.pdf> Consulta: [04/06/2015].

