

Aspectos prácticos para proyectos de datos abiertos en las administraciones públicas

Juan-Antonio Pastor-Sánchez

31 enero 2014

Pastor-Sánchez, Juan-Antonio (2014). "Aspectos prácticos para proyectos de datos abiertos en las administraciones públicas". *Anuario ThinkEPI*, v. 8, pp. 313-317.



Resumen: La publicación de datos abiertos gestionados por las administraciones públicas comienza a considerarse una acción estratégica para el desarrollo social y económico. El apoyo legislativo y normativo sobre este tema debe complementarse con consideraciones prácticas relacionadas con aspectos técnicos concretos. Estos aspectos afectan a los metadatos descriptivos, que contienen catálogos especializados, sobre conjuntos de datos, el uso de formatos abiertos y la necesidad de contemplar los proyectos de datos abiertos en el ciclo de información de las organizaciones. Los datos abiertos son una fuente de innovación y nuevos modelos de negocio pero precisan de mecanismos eficientes de localización de conjuntos de datos y herramientas para su acceso y la reutilización.

Palabras clave: Datos abiertos, Datos abiertos en el sector público, Interoperabilidad, Reutilización de la información, Formatos abiertos.

Title: Practical aspects of open government data projects

Abstract: Publishing open data managed by public administrations is beginning to be considered a strategic action for social and economic development. The legislative and policy support on this issue should be complemented by practical considerations related to certain technical aspects. These aspects have an impact on the descriptive metadata datasets of specialized catalogs, the use of open formats and the need to consider open data projects in the cycle of information in organizations. Open data is a source of innovation and new business models that requires efficient mechanisms for locating datasets and tools to access and reuse them.

Keywords: Open data, Open government data, Interoperability, Reuse of information, Open formats.

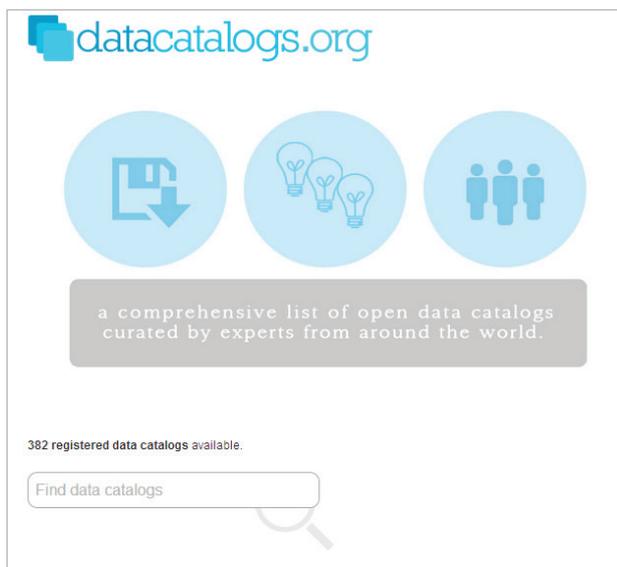
1. Introducción

La *Directiva Europea 2003/98/CE* para la reutilización de la información en el sector público, planteó la conveniencia de la disponibilidad de los datos de carácter público elaborados o gestionados por las administraciones. Este planteamiento propone un medio para que el conjunto de la sociedad participe en la vigilancia del funcionamiento de los servicios públicos (Yu; Robinson, 2012). Para las administraciones públicas supone una mayor transparencia en el uso de los recursos públicos, al tiempo que se mejora su capacidad de interconexión y la reutilización de información. El sector privado también se beneficia al tener acceso a gran cantidad de datos, convenientemente licenciados y que pueden servir como materia prima para el desarrollo de nuevos productos y servicios. En general, los datos abiertos pueden ser el catalizador para generar innovación, nuevos modelos de negocio y valores, que podrían

generar riqueza económica y social (Prince; Jolías, 2013).

La directiva europea se transpone a España mediante la *Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público* y la puesta en marcha del *Proyecto Aporta*. Desde entonces, en España se ha trabajado de forma intensa en este campo. Son numerosos los proyectos y servicios de datos abiertos existentes actualmente (Ferrer-Sapena; Peset; Aleixandre-Benavent, 2011), hasta el punto de que nuestro país se ha convertido en un referente en esta materia. Otra muestra de ello es la puesta en marcha del capítulo español de *OKFN (Open Knowledge Foundation)* que conforma un grupo de interés muy activo en la materia.

La nueva *Directiva 2013/37/UE* refuerza el papel de los datos abiertos en la reutilización de la información pública, incrementando el nivel de compromiso y de exigencia de la Administración a este respecto. Por su parte, las normas técni-



<http://datacatalogs.org>

cas sobre interoperabilidad publicadas en 2012 desarrollan aspectos concretos del *Real Decreto 4/2010, sobre el Esquema nacional de interoperabilidad en el ámbito de la administración electrónica*.

Puesto que existe una base normativa suficiente y numerosas iniciativas para la publicación de datos abiertos, resulta conveniente reflexionar sobre ciertos aspectos prácticos que faciliten la sostenibilidad de estos proyectos.

“La disponibilidad de grandes cantidades de datos abiertos redundará en mayores expectativas de creación de aplicaciones dirigidas al sector privado y al usuario final”

2. El compromiso *open data*

Sin duda el concepto de datos abiertos en el ámbito de la Administración pública resulta el más atractivo. Existen más de 350 catálogos de datos registrados en *datacatalogs.org* vinculados al concepto de *open government data*. Cada catálogo puede llegar a describir y referenciar cientos o miles de conjuntos de datos. Las directivas europeas indican que el desarrollo de nuevos servicios y productos precisan datos “en crudo”. Por lo tanto, la disponibilidad de grandes cantidades de datos abiertos redundará en mayores expectativas de creación de aplicaciones dirigidas al sector privado y al usuario final.

Los datos abiertos deben ofrecerse lo suficientemente desagregados o detallados como para que su explotación resulte útil para elaborar

aplicaciones o servicios, o para su uso con fines de escrutinio de la actividad de las administraciones. Es necesario que los datos se encuentren actualizados para evitar cualquier incertidumbre en cuanto a su fiabilidad durante su reutilización.

Las organizaciones han de ser conscientes de que la publicación de datos abiertos también implica aceptar el compromiso de mantenerlos accesibles y actualizados a lo largo del tiempo y con los máximos niveles de detalle, calidad y fiabilidad posibles.

El compromiso *open data* no se limita a los propios datos, también debe aplicarse a los catálogos de datos abiertos, que generalmente operan mediante autoregistro. Muchos recursos suelen cambiar su ubicación e incluso desaparecen, y pese a ello su huella permanece en los catálogos de datos. En otras ocasiones los conjuntos de datos se describen incorrectamente e incluso a veces los registros del catálogo ni siquiera hacen referencia a datos en sí, sino a direcciones genéricas de sitios web o buscadores en bases de datos.

“Las organizaciones que gestionan los catálogos deben definir mecanismos para asegurar la calidad de la información de sus registros”

Teniendo en cuenta que los catálogos constituyen un punto de acceso a los datos abiertos, resulta esencial que los metadatos que los describen sean verificados y actualizados con frecuencia (Zuiderwijk; Jeffery; Janssen, 2012). Del mismo modo que los responsables de la publicación de datos abiertos deben adquirir el compromiso de su mantenimiento, las organizaciones que gestionan los catálogos deben definir mecanismos para asegurar la calidad de la información de sus registros.

3. El formato y la estructura sí importan

Una de las bases de la disponibilidad de conjuntos de datos abiertos se centra en el uso de formatos abiertos. El *Real Decreto 4/2010, sobre el Esquema nacional de interoperabilidad en el ámbito de la administración electrónica*, indica de forma expresa que la Administración pública española debe usar estándares abiertos, y de forma complementaria, estándares de uso generalizado por los ciudadanos. De este modo los diferentes usuarios y desarrolladores tendrían la libertad de escoger uno u otro formato para obtener los datos, sin tener que atarse a un *software* o a una solución tecnológica específica. Por este moti-

vo resulta paradójico comprobar en los catálogos de datos abiertos el uso masivo de formatos propietarios fuertemente ligados al uso de soluciones específicas de software.

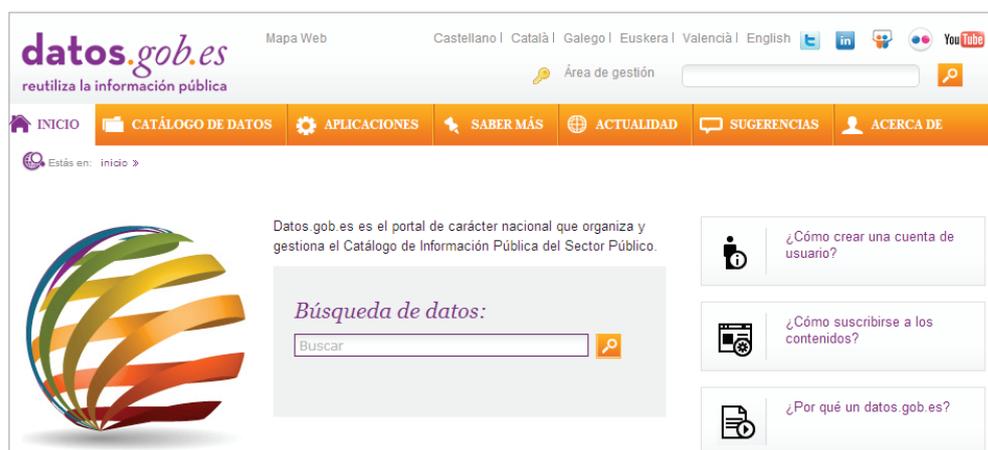
Profundizando en este aspecto la resolución de 3 de octubre de 2012 (BOE 31-10-2012), que aprueba la norma técnica sobre interoperabilidad de catálogo de estándares, establece los estándares y los organiza en función de una cadena de interoperabilidad formada por los siguientes eslabones:

- accesibilidad multicanal, integrada y segura;
- infraestructuras y servicios asociados;
- integración de sistemas y servicios;
- modelos e integración de datos.

Debe tenerse en cuenta que una reutilización eficaz de los datos parte de la representación de la información mediante principios de interoperabilidad semántica (Davies et al., 2008) que se relaciona directamente con los estándares de modelos e integración de datos. Sin embargo, muchos de los conjuntos de datos abiertos que ofrece la Administración pública en España utilizan formatos cuya función se encuadra en el acceso a los datos. Es muy común el uso de documentos pdf, html o de procesadores de texto para publicar datos estructurados, lo cual hace imposible su reutilización de un modo ágil y sencillo.

Un breve análisis de catálogos como *Datos.gob.es*, *Open data Euskadi* y *Datos abiertos genca*, indica que de los más de 5.200 conjuntos de datos registrados en estos tres catálogos, únicamente unos 150 utilizan alguno de los formatos de rdf. Esto indica que, aunque se habla mucho de datos abiertos en el seno de la Administración pública, todavía queda un largo camino por recorrer para alcanzar el ideal de la interoperabilidad semántica. Afortunadamente, la *Guía de aplicación de la norma técnica de interoperabilidad de reutilización de recursos de información (MHAP, 2013)* constituye un valioso documento de referencia que puede ayudar en esta tarea.

Un problema diferente es el que deben afrontar los editores al escoger qué vocabularios, esquemas de metadatos u ontologías utilizarán para estructurar y representar conjuntos de datos abiertos. Con respecto al contenido de los propios catálogos de datos la solución a este problema



<http://datos.gob.es>

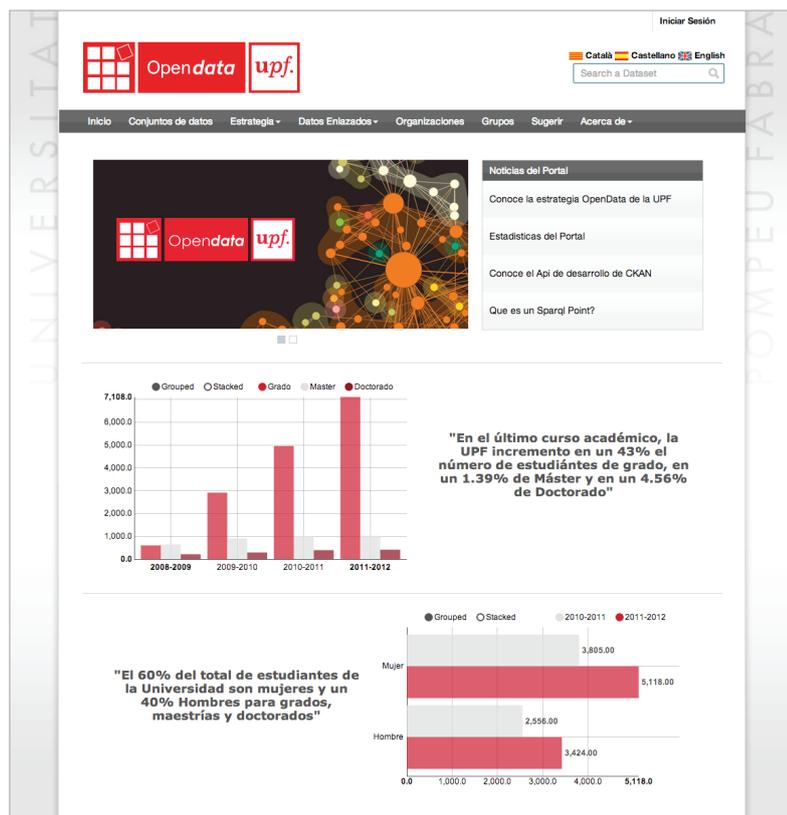
está clara: el vocabulario DCAT (Maali; Erickson, 2014) ha de utilizarse para este cometido, tal y como establece la norma técnica sobre interoperabilidad del catálogo de estándares.

Sin embargo no existe una solución única para los propios conjuntos de datos. El abanico de vocabularios rdf es muy amplio, e incluso en algunos casos es posible que sea preciso definir alguno nuevo. Esto implica tomar decisiones, en ocasiones arriesgadas, acerca de la correspondencia y el nivel de agregación de los datos primarios, así como su equivalencia con respecto a los elementos de un vocabulario. Puesto que las administraciones públicas hacen un uso intensivo de sistemas de bases de datos relacionales, una posible solución a este problema sería la aplicación de la recomendación del W3C para el mapeado directo de datos relacionales a rdf (Arenas et al., 2012).

“Una adecuada política informativa debe identificar qué datos son susceptibles de publicarse en abierto, así como el modo correcto en el que debe hacerse”

4. Visión integral del ciclo de los datos abiertos

Desde el punto de vista técnico, un conjunto de datos abiertos tiene un ciclo de vida que comprende la extracción de los datos, su almacenamiento, revisión, interconexión con otros datos abiertos, clasificación y mantenimiento (Hyland; Atemezing; Villazón-Terrazas, 2013). Sin embargo también hay que contemplar dicho ciclo de vida desde la dimensión corporativa de las organizaciones que los publican.



<http://data.upf.edu/es/main>

Por lo tanto hay que tener una visión sobre los datos abiertos como un elemento ligado a los procesos de gestión de información. Una adecuada política informativa en cualquier Administración pública debe incluir una serie de tareas de auditoría de información, que permita identificar qué datos son susceptibles de publicarse en abierto, así como el modo correcto en el que debe hacerse.

“No es conveniente poner en marcha proyectos de datos abiertos sin considerar su sostenibilidad a lo largo del tiempo”

A este respecto, merece la pena destacar el trabajo realizado por la *Universitat Pompeu Fabra*. No solamente ofrece una serie de datos abiertos organizados a través de un portal con su correspondiente catálogo, sino que también refleja un considerable esfuerzo para comprender el ecosistema informativo de la organización en el que se desenvuelve un proyecto de datos abiertos (**Pantoja, 2013**).

La localización y análisis de fuentes de datos corporativas resultan vitales para comprender dicho ecosistema y definir las estructuras y niveles de detalle para establecer la oferta de datos abier-

tos de una organización. La delimitación precisa de los objetivos que se persiguen al desarrollar portales de datos abiertos permite diseñar un ciclo eficiente de los datos abiertos, integrándose en el resto de procesos de gestión corporativos.

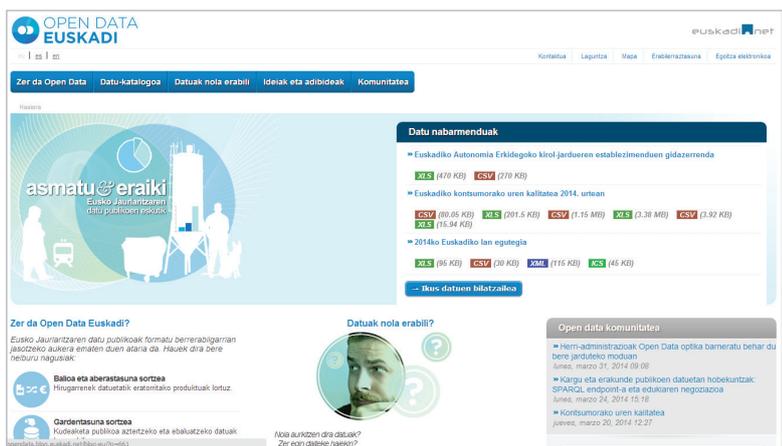
5. Conclusiones

Los datos abiertos suponen una fuente de innovación y nuevos modelos de negocio y perfiles profesionales. Sin embargo, también presentan una serie de retos que van más allá de la aplicación de tecnologías y que precisan reflexionar y tomar decisiones. No es conveniente poner en marcha proyectos de datos abiertos sin considerar su sostenibilidad a lo largo del tiempo. Hay que analizar cuidadosamente la calidad y cantidad de los datos, la utilidad y el interés que pueda tener su explotación, su frecuencia de actualización, los formatos a utilizar, el grado de reutilización e interoperabilidad semántica, las fuentes corporativas, su incardinación en la estructura

informativa y de gestión de la organización, etc. Un proyecto de datos abiertos puede fracasar si la organización que los publica no está preparada para su mantenimiento y evolución.

Cierto es que la publicación de datos abiertos concierne directamente a las administraciones públicas. Pero no es menos cierto que muchas empresas y organizaciones privadas también podrían plantearse estas iniciativas. Para el sector público se trata de una obligación que poco a poco está creando una demanda social y que favorece el intercambio eficiente de información entre administraciones. En el caso del sector privado podría aplicarse para ofrecer datos sobre productos, servicios y actividades, al tiempo que formaría parte de la estrategia de transparencia y de responsabilidad social corporativa.

Desde la perspectiva más técnica deberían desarrollarse propuestas que ayuden a la localización, estructuración y aprovechamiento de los datos. La dispersión y el entramado de catálogos deberían tender a la definición de una política pública al respecto. El registro normativo en un catálogo nacional o territorial de conjuntos de datos abiertos de la Administración pública es compatible con su integración en sedes electrónicas o la creación de portales o catálogos de datos abiertos de ayuntamientos, universidades, comunidades autónomas u otros organismos. La neutralidad de representación rdf es otro aspecto



<http://opendata.euskadi.net/w79-home/es>

que merece la consideración cuando se precise la publicación de datos en su forma primaria y completa, libre de la interpretación que puede suponer la aplicación de vocabularios u ontologías específicas.

“El éxito se encuentra en la sencillez de su localización y acceso, así como en las aplicaciones y servicios que se desarrollen a partir de su explotación”

También debe considerarse la creación de APIs públicas (como la que nos ofrece la agenda de eventos de *KulturKlic*), que permitan seleccionar y recuperar datos específicos. No hay que olvidar que el éxito real que en un futuro tengan los datos abiertos no se encuentra en sus implicaciones políticas o tecnológicas, sino en la sencillez de su localización y acceso, así como en las aplicaciones y servicios que se desarrollen a partir de su explotación.

6. Recursos web mencionados

Portal Datos.gov.es
<http://datos.gov.es>

Datacatalogs
<http://datacatalogs.org>

Open data Euskadi
<http://opendata.euskadi.net>

Datos abiertos Gencat
<http://www20.gencat.cat/portals/site/dadesobertes>

UPF open data
<http://data.upf.edu/les/main>

API de la agenda de eventos de Kulturklic
<http://www.kulturklic.euskadi.net/lang/es/llaguntza-ayudalapi>

7. Referencias

Arenas, Marcelo; Bertails, Alexandre; Prud'hommeaux, Eric; Sequeda, Juan (2012). *A direct mapping of relational data to RDF. W3C Recommendation. W3C*, 27 septiembre.
<http://www.w3.org/TR/2012/REC-rdb-direct-mapping-20120927>

Davies, Jim; Harris, Steve; Crichton, Charles; Shukla, Aadya; Gibbons, Jeremy. "Metadata standards for semantic interoperability in electronic government". En: *Proceedings of the 2nd international conference on Theory and practice of electronic governance ICEGOV '08*, pp. 67-75.
<http://www.cs.ox.ac.uk/people/jeremy.gibbons/publications/metadata-egov.pdf>

Ferrer-Sapena, Antonia; Peset, Fernanda; Aleixandre-Benavent, Rafael (2011). "Acceso a los datos públicos y su reutilización open data y open government". *El profesional de la información*, v. 20, n. 3, pp. 260-269.
<http://eprints.rclis.org/20997>
<http://dx.doi.org/10.3145/epi.2011.may.03>

Hyland, Bernadette; Atemezeng, Ghislain; Villazón-Terrazas, Boris (2013). *Best practices for publishing linked data. W3C Working Group Note. W3C*, Versión actualizada en 2014.
<http://www.w3.org/TR/ld-bp>

Maali, Fadi; Erickson, John (2014). *Data catalog vocabulary (DCAT). W3C Recommendation. W3C*, 16 enero.
<http://www.w3.org/TR/2014/REC-vocab-dcat-20140116>

MHAP (2013). *Guía de aplicación de la Norma técnica de interoperabilidad de reutilización de recursos de información*. Ministerio de Hacienda y Administraciones Públicas.
<http://administracionelectronica.gob.es>

Pantoja, Jorge (2013). *Linked open data at the UPF*. Barcelona: Universitat Pompeu Fabra (Tesis de master).
<http://data.upf.edu/upf/docs/2013/jorgepmlodatumupf.pdf>

Prince, Alejandro; Jolías, Lucas (2013). "Las fuentes conceptuales del gobierno abierto". *Cuadernos de comunicación e innovación*, n. 94, pp. 48-57.
http://telos.fundaciontelefonica.com/seccion=1266&idioma=es_ES&id=2013021317570001&activo=6.do

Yu, Harlan; Robinson, David G. (2012). "The new ambiguity of Open Government". *UCLA Law Review Disc*, 178.
<http://dx.doi.org/10.2139/ssrn.2012489>

Zuiderwijk, Anneke; Jeffery, Keith; Janssen, Marijn (2012). "The necessity of metadata for linked open data and its contribution to policy analyses". En: *Proceedings of the International Conference for E-Democracy and Open Government CeDEM12*, pp. 281-294.