

Carencias informativas de los datos abiertos en España

Fernanda Peset y Antonio-Lázaro Fernández-López

6 diciembre 2013

Peset, Fernanda; Fernández-López, Antonio-Lázaro (2014). "Carencias informativas de los datos abiertos en España". *Anuario ThinkEPI*, v. 8, pp. 318-321.



Resumen: Se analiza la apertura de datos de estadística pública en los catálogos de datos de los gobiernos central y autonómicos españoles. Se parte de la hipótesis de que su tratamiento documental es insuficiente, si tenemos en cuenta que los datos estadísticos son especialmente relevantes para *open data*. Se estudian tres cuestiones con respecto a la estadística oficial: que cumpla con los requerimientos que se solicitan a los datos abiertos, la posibilidad de ser encontrada en los catálogos y que pueda ser correctamente comprendida de cara a la reutilización. Estos aspectos implican definir qué es un dato abierto y qué es estadística pública; estudiar si en los portales es posible identificar esa

información estadística; y si el tratamiento documental permitiría una reutilización con garantías. Los resultados arrojan una gran disparidad, tanto en la forma de denominar la información estadística (encontrabilidad) como en la manera de ofrecerla (accesibilidad). Por tanto los portales aumentan la dispersión a la hora de recuperarla y sólo algunos describen los conjuntos de datos de manera que resulten inteligibles por sí mismos. Se concluye la necesidad de concentrar esfuerzos o al menos de coordinar las acciones con respecto a la oferta de datos abiertos en España.

Palabras clave: Información estadística, Actividad estadística pública, Catálogos de datos, Datos abiertos, Encontrabilidad, Accesibilidad, España.

Title: Information gaps in open data in Spain

Abstract: This paper studies how public statistical data are becoming open in the catalogs of Spain's central and regional governments. It starts with the hypothesis that documentary treatment is insufficient, especially if we consider that statistics are particularly relevant in the open data world. This paper examines three issues regarding official statistics: what meets the requirements required of open data, how data can be found in the catalogs, and whether they can be properly understood ahead of reuse. These three aspects involvedefining open data and public statistics, considering whether it is possible to identify the statistical information in the portals, and finally determining if the documentary treatment allows reuse with guarantees. The results show a wide disparity, both in the form of designating the statistical information (findability) and how to offer it (accessibility). Therefore, a portal can increase data dispersion when they are recovered. Furthermore, only some portals describe datasets so that they are readily intelligible. We conclude with the need to concentrate efforts or at least coordinate actions with respect to the supply of open data in Spain.

Keywords: Statistical information, Official statistic activity, Data catalogs, Open Data, Findability, Accessibility, Spain.

1. Introducción

La evolución hacia la apertura de los datos (*open data*), es ya imparable. Las instancias políticas (Kroes, 2013), las administraciones (*Ley de reutilización*, 2007) y hasta los sectores académicos (Ferrer-Sapena; Sánchez-Pérez, 2012; Hernández-Pérez; García-Moreno, 2013; Dotor-Alonso, 2013) están hablando de ello. Como

siempre que están presentes tantos actores en un escenario, un mismo término tiene significados diferentes para cada uno de ellos. En este texto veremos aspectos derivados de los datos públicos abiertos (*ODG open data government*), uno de cuyos hitos más conocidos son las declaraciones de Obama en 2009 sobre apertura de la Administración.

Reflexionamos sobre cómo se han implemen-

tado los datos procedentes de la actividad estadística pública en los catálogos de datos de los gobiernos autonómicos, dado que esta información es frecuente en las aplicaciones que consumen datos abiertos. A pesar de esta intensa presencia consideramos que su tratamiento informativo o documental muestra insuficiencias por los diferentes niveles de análisis y otros aspectos. En este texto se estudian tres cuestiones con respecto a la estadística oficial:

- que cumpla con los requerimientos que se solicitan a los datos abiertos,
- la posibilidad de ser encontrada en los catálogos,
- que pueda ser correctamente comprendida de cara a la reutilización.

Hemos buscado en cada uno de los portales de gobierno abierto:

- si es posible identificar la información estadística pública, puesto que, como ya se ha comentado, la información numérica es uno de los pilares de la recombinación de datos;
- si el tratamiento informativo que se realiza permite, a nuestro entender, una reutilización con garantías.

2. ¿Qué es?

El primer aspecto es fijar lo que entendemos por información estadística pública y por *open data*. Dado que no existe una definición de estadística oficial que contemple todas las situaciones del panorama español, asumiremos que ésta es la producida por la actividad estadística de los organismos públicos con competencias en esta materia. Actividades estadísticas son las “que conducen a la obtención, la recopilación, la elaboración y la ordenación sistemática de datos...” (*Ley de Cantabria*, art.2, 1990).

Acordar qué son *open data* también resulta complicado. Recordemos que en numerosas ocasiones los concebimos como datos numéricos y gran parte de ellos son producto de la actividad estadística oficial, cuyos *datasets* también son llamados microdatos. La *Ley de reutilización* (2007), por ejemplo, indica que la información del sector público –a la que llama documentos- ha de estar disponible en la Red en unos formatos determinados y tener una declaración de lo que puede y no puede hacerse para que pueda reutilizarse.

Por su parte, la *OKF (Open Knowledge Foundation)* dice: “a piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike”.

<http://opendefinition.org>

Desde el sector de la gestión de la información consideramos que han de tener los suficientes

metadatos para que sean interpretables y se reutilicen con garantías de fiabilidad. Quizá por ello la definición del informe de la *Royal Society* (2012) sobre datos abiertos, es a nuestro entender también adecuada para los *open data* de la estadística oficial. Los define como datos que cumplen ciertos criterios cuando son liberados -también llamada apertura inteligente/competitiva-, por lo que deben ser:

- accesibles (localizables): depositados de manera que puedan ser encontrados, y en una forma en que puedan ser inmediatamente usados;
- certificados (acreditados): que se ha realizado una evaluación sobre la credibilidad de los datos;
- inteligibles (interpretables): que puedan ser entendidos por quienes desean conocer algo;
- reutilizables, o estar en un formato por el que otros puedan usar esta información; y por tanto, requieren la información necesaria para ello, los metadatos.

“Los datos abiertos han de ser accesibles, certificados, inteligibles y reutilizables”

3. ¿Se encuentran?

La información estadística ya estaba disponible desde hace años en internet, si bien no es demasiado fácil encontrar exactamente la que deseamos para un problema determinado (**Fernández-López**, 2013). Además, nuestro país cuenta con una estructura muy característica al tener diferentes niveles: un sistema para la Administración General del Estado y varios autonómicos.

Hacer disponible la información pública es uno de los pilares de la transparencia de la Administración (España, 2013). Pero ¿dónde están esos conjuntos de datos en los que pensamos los usuarios potenciales?, ¿es posible distinguir la información estadística entre las funciones de los catálogos de datos abiertos?

Los portales de gobierno abierto que han surgido (tanto el central como los autonómicos) están ofreciendo todo tipo de información abierta. A partir de 2005, con el *Plan Avanza* y el proyecto *Aporta*, el gobierno de España ha procurado ordenar el panorama sobre los datos abiertos, creando el portal *datos.gob.es* en 2011. Los catálogos actúan como primer acceso único a los diferentes recursos web que pueden o no estar identificados como proveedores de datos abiertos.

<http://datos.gob.es>

<http://datos.gob.es/datos/?q=node/521>

Ahora bien, el estatal no es el único portal que existe, tal y como se observa en su inventario de iniciativas españolas, que nosotros hemos completado con otros portales autonómicos abiertos. Esta diversidad de catálogos se convierte en sí misma en una dificultad para encontrar los conjuntos de datos.

<http://datos.gob.es/datos/?q=node/237>

“Se aprecia una gran disparidad en la forma en que los portales ofrecen los metadatos”

Además, hemos analizado el tratamiento que dan a la información estadística pública con el objetivo de saber si se identifica claramente y se puede recuperar de forma sistemática.

Los portales más exhaustivos ofrecen un apartado específico para este tipo de información, como en el vasco, que tiene un apartado “Buscar datos estadísticos”. Sin embargo, en otros son recuperables como “Tipo de datos” o como una etiqueta en una nube de tags. Incluso existen casos en que ni siquiera podría llegarse a esta información utilizando el buscador, por ejemplo, si un censo o una encuesta no se etiquetaran también como estadística.

4. Metadatos

Como último paso, hemos determinado si esa oferta de datos se acompaña de los metadatos específicos de los conjuntos para hacerlos comprensibles y por tanto reutilizables (inteligibles en palabras de la *Royal Society*). Los metadatos en el ámbito de *Open data* deben informar de los *datasets* a los que hacen referencia y tienen que poder ser recuperados al tiempo que el *dataset* concreto, tal y como indica la *Ley Foral 11/2012 de la transparencia y del gobierno abierto de Navarra*.

Para ello hemos buscado en los catálogos un conjunto de datos

con información estadística acerca del mercado de trabajo, específicamente la tasa de actividad o el número de población inactiva, ya que esos datos son utilizados por todos los órganos estadísticos de las Comunidades autónomas para obtener sus propios resultados. Esta información se obtiene de la operación estadística coyuntural *Encuesta de Población Activa (EPA)*, del *Instituto Nacional de Estadística*. De este conjunto de datos nos preguntamos si podemos encontrar la “Tasa de actividad” o “Inactivos” en el portal de datos abiertos. Y en relación con la metainformación necesaria, si se ofrece información para la reutilización, como por ejemplo, ¿qué es la “tasa de actividad” o ser “inactivo”?

En términos generales se aprecia de nuevo una gran disparidad en la forma en que los portales ofrecen los metadatos. En el mejor de los casos los asocian al *dataset* pero en otros muchos, sin embargo, los conjuntos de datos no se acompañan de sus metadatos específicos sino de los que informan de la operación estadística en su totalidad. Por tanto complican en exceso su interpretación y su reutilización directa. Algunos *datasets*, incluso, no cuentan con ningún metadato.

Por ejemplo, el catálogo de *datos.gob*, que ofrece acceso indirecto a la información estadística pública, descarga la responsabilidad de ofrecer metadatos específicos a las páginas a las que redirige. Los metadatos de sus registros no llegan al nivel de detalle que requeriría la descripción de conjuntos de datos estadísticos. En ocasiones

Comunidad	Iniciativa
Catalunya	<i>Dades obertes</i> http://www20.gencat.cat/portal/site/dadesobertes
Extremadura	<i>Proyecto reutiliz@</i> Error permanente en http://www.extremadurareutiliza.es
Islas Baleares	<i>Dades obertes CAIB</i> http://www.caib.es/caibdatafront/index?lang=es
Navarra	<i>Open data Navarra</i> http://gobiernoabierto.navarra.es/es/open-data
Principado de Asturias	Datos de Asturias http://risp.asturias.es/catalogo/index.html Enlace no funciona
Rioja	<i>Dato abierto Rioja</i> http://www.larioja.org/npRioja/default/index.jsp?idtab=758984
Aragón	<i>Aragón open data</i> http://opendata.aragon.es
Galicia	<i>Abert@s Portal open data de la Xunta de Galicia</i> http://abertos.xunta.es/portada
País Vasco	<i>Open data Euskadi</i> https://euskadi.net/w79-home/es
Castilla y León	<i>Portal de datos abiertos</i> http://www.datosabiertos.jcyl.es
Canarias	La dirección en el catálogo de <i>gob.es</i> es la del <i>Istac</i>
Castilla la Mancha	<i>Portal de datos abiertos de JCCM</i> http://opendata.jccm.es
Andalucía	Junta de Andalucía. Datos abiertos http://www.juntadeandalucia.es/datosabiertos/portal.html

enlaza a bases de datos mayores como puede ser *Inbase*, donde el usuario ha de utilizar formularios de consulta para llegar a la información que desea. Y, por supuesto, los metadatos están aparte, en documentos extensos asociados al conjunto de operaciones estadísticas.

En cuanto a los portales autonómicos se aprecia que algunos también efectúan reenvíos a los institutos de estadística (Andalucía o Galicia). Otros sin embargo sirven la información desde sus páginas, algunos con metadatos (Euskadi, Navarra o Catalunya). Pero la mayoría no lo hace: Baleares, Rioja o Castilla-León. Incluso algunos ni siquiera funcionan, a la luz de nuestros repetidos intentos hasta el 4 de diciembre de 2013.

“Los contenidos han de ser la clave de este tipo de productos de información, no la tecnología ni el diseño”

5. Conclusiones

Observamos cómo hay sensibilidades diferentes a la hora de abordar la incorporación de datos de la estadística oficial en los catálogos de los portales de datos abiertos. Encontramos diferentes modelos, como los catálogos que funcionan de manera referencial y remiten a los órganos estadísticos de sus respectivas administraciones, descargando en ellos la responsabilidad de ofrecer la metainformación y los formatos adecuados.

Con tal variedad de portales y de formas de distribuir la información, en lugar de facilitar que se encuentren los conjuntos de datos, se han añadido dificultades. Al ser tantos se pierden las ventajas de una economía de escala: pocos portales bien diseñados que incluyeran una masa crítica suficiente de *datasets*. Nos encontramos con un problema conocido ya por nosotros: muchas formas de entrada a pocos contenidos, que se duplican una y otra vez, quizá debido a la facilidad para compartir recursos gracias a la interoperabilidad técnica.

Esta misma dispersión de recursos impide saber qué información estadística vamos a encontrar y cómo la vamos a encontrar en un catálogo:

- si se recoge información de cualquier nivel administrativo dentro del área geográfica o sólo del nivel autonómico;
- si son reenvíos a otras páginas;
- si se van a describir los conjuntos de datos, o al menos los más significativos.

Por otra parte, en cada uno de los portales la estadística oficial se recupera de las formas más variadas. Eso sin entrar en que la multiplicidad de recursos produce más errores, como encontrar estadística de una autonomía cuando se busca la

estatal. Y estas conclusiones podrían dar lugar a interpretarnos de forma errónea. Tampoco creemos que un solo portal de tipo centralizado, como *datos.gob.es*, sea la solución.

Con esta reflexión queremos hacer notar lo que conocemos como gestores de información: los contenidos han de ser la clave de este tipo de productos de información, no la tecnología ni el diseño. Eso significa que por métodos manuales o automatizados deben asignarse los metadatos que necesita un *dataset* para ser reutilizado de forma fiable. Y en esto de nuevo nuestra profesión mucho que decir y un largo camino por andar.

6. Bibliografía

Cantabria (1990). “Ley 3/1990, de 21 de marzo, de Estadística de Cantabria”. *BOC*, 9 de abril.

http://www2.parlamento-cantabria.es/UserFiles/File/leyes/III%20Legislatura/LEY3_90.pdf

Dotor-Alonso, María (2013). *Datos abiertos de seguridad ciudadana*. Valencia: Universidad Politécnica de Valencia [Tesis final de máster].

España (2013). “Ley 19/2013, de 9 de diciembre. Ley de transparencia, acceso a la información pública y buen gobierno”. *BOE*, 10 diciembre.

<http://www.leydetransparencia.gob.es/anteproyecto/index.htm>

<https://www.boe.es/boe/dias/2013/12/10/pdfs/BOE-A-2013-12887.pdf>

España (2007). “Ley 37/2007, de 16 de noviembre, sobre reutilización de la información del sector público”. *BOE*, 17 noviembre, n. 276.

<https://www.boe.es/buscar/doc.php?id=BOE-A-2007-19814>

Fernández-López, Antonio-Lázaro (2013). *Análisis y diagnóstico de la difusión de la información y documentación estadística de las administraciones públicas en España*. Valencia: Universidad Politécnica de Valencia [Trabajo de DEA].

Ferrer-Sapena, Antonia; Sánchez-Pérez, Enrique A. (2013). “Open data, big data: ¿hacia dónde nos dirigimos?”. *Anuario ThinkEPI*, v. 7, pp. 150-156.

Hernández-Pérez, Tony; García-Moreno, María-Antonia (2013). “Datos abiertos y repositorios de datos: nuevo reto para los bibliotecarios”. *El profesional de la información*, v. 22, n. 3, pp. 259-263.

http://eprints.rclis.org/19524/1/postprint_Hernandez_Garcia.pdf

<http://dx.doi.org/10.3145/epi.2013.may.10>

Kroes, Neelie (2013). *Opening up scientific data*. European Union. Speech/13/236, 18 marzo.

http://europa.eu/rapid/press-release_SPEECH-13-236_en.htm

Red.es. *Metadatos para el catálogo datos.gob.es*.

http://datos.gob.es/datos/sites/default/files/files/5_met_06.pdf

The Royal Society (2012). *Science as an open enterprise*. The Royal Society Science Policy Centre report. ISBN: 978 0 85403 962 3

<http://royalsociety.org/policy/projects/science-public-enterprise/report>