

El Portal de datos abiertos de la Unión Europea: análisis y evaluación

Rosario ARQUERO AVILÉS

Universidad Complutense de Madrid. Biblioteconomía y Documentación
carquero@ucm.es

Gonzalo MARCO CUENCA

Universidad Complutense de Madrid. Biblioteconomía y Documentación
gmarco@ucm.es

Recibido: Mayo 2014

Aceptado: Junio 2014

Resumen: El movimiento de *datos públicos abiertos* (*Open Data Government*) se vincula a los principios de transparencia, participación y colaboración. En línea con estos principios, las políticas de apertura y liberación de datos de los gobiernos entroncan con la faceta de la puesta a disposición de entidades, empresas y ciudadanos de información del sector público para que pueda ser reutilizada y añadir valor. En este contexto, se presenta un análisis del Portal de datos abiertos de la Unión Europea como fuente de datos y punto de acceso único a los datos elaborados por instituciones y otros organismos de la Unión Europea. El objetivo del trabajo se centra en abordar el estudio de los contenidos del mismo, los estándares en los que se basa, así como los principales conjuntos de datos, formatos y aplicaciones que ofrece.

Palabras clave: Datos abiertos; Portal de datos abiertos de la Unión Europea; Gobierno abierto; Reutilización de la información del sector público; Portales de datos abiertos; Formatos abiertos

European Union Open Data Portal: analysis and evaluation

Abstract: The open data movement (Open Government Data) is linked to the principles of transparency, participation and collaboration. In line with these principles, the policies of opening and release of data from governments connect with the aspect of the provision of organizations, companies and citizens to public sector information that can be reused and add value.

In this context is presented an analysis of the open data portal of the European Union as a data source and a single point of access to the data produced by other institutions and bodies of the European Union. The aim of the work is focused on addressing the study of its contents, standards on which it is based, as well as major datasets, formats and applications available.

Keywords: Open data; European Union Open Data Portal; Open government; Re-use Information of Public Sector; Open data portals; Open formats

1 INTRODUCCIÓN

Las iniciativas de datos abiertos se constituyen en un elemento clave que favorece las políticas de transparencia de las Administraciones Públicas, aseguran el libre acceso a los datos a todos los ciudadanos y promueven oportunidades de negocio a partir de la reutilización de los datos públicos como materia prima.

Estas iniciativas, ampliamente extendidas durante los últimos años, se han apoyado para su impulso en dos pilares clave: el *desarrollo de las nuevas tecnologías*, que facilitan que la información pueda ser publicada, relacionada y compartida de forma eficaz, y las *políticas de gobierno abierto u open government*, que fomentan estrategias orientadas al avance de una sociedad digital asentada sobre los principios de transparencia, colaboración y participación.

Estos principios son formulados por primera vez por el Presidente de los Estados Unidos, Barack Obama, en enero de 2009 en *el Memorandum de Transparencia y Gobierno Abierto* (Obama, 2009: 4685-4686). Con tal comunicado la administración de este país adquiriría un compromiso con el gobierno abierto y con la aplicación de acciones específicas para el desarrollo de los principios enunciados. El gobierno federal de los Estados Unidos responde rápidamente al llamamiento con la inauguración de su portal de datos abiertos en abril de ese mismo año¹, constituyéndose como uno de los portales pioneros en la distribución de datos abiertos de una administración pública. Al poco tiempo, otros países como Reino Unido, Nueva Zelanda o Canadá seguirán esta tendencia de apertura de datos a través de sus portales.

La existencia de liderazgo y voluntad política es, sin duda alguna, otro de los puntos fuertes para el progreso de esta filosofía alrededor del gobierno abierto. En el caso de Europa, esta voluntad y liderazgo queda expresada en las numerosas intervenciones de Neelie Kroes, Vicepresidenta de la Comisión Europea y actual Comisaria de la Agenda Digital Europea, autodeclarada “fan” del movimiento de datos abiertos y partidaria de una información pública para todos (Kroes, 2011).

Las políticas europeas en torno a la reutilización de la información pública suponen también un importante empuje para el desarrollo de los portales de datos abiertos. Estas políticas quedan reflejadas en diferentes documentos publicados por la Unión Europea a lo largo de varios años, como son: el *Libro Verde sobre la información del Sector Público en la Sociedad de la Información* (Comisión Europea, 1998), la *Directiva 2003/98/CE* relativa a la reutilización de la información del sector público (Unión Europea, 2003), la comunicación *COM (2011)882* de la Comisión sobre Datos abiertos: un motor para la innovación, crecimiento y gobernanza transparente (Comisión Europea, 2011), la Decisión

¹ El portal de datos abiertos de Estados Unidos puede consultarse en la dirección: <http://www.data.gov/>

2011/833/UE de la Comisión sobre la reutilización de los documentos de la Comisión (Unión Europea, 2011) o la reciente Directiva 2013/37/UE por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información en el sector público (Unión Europea, 2013).

La Directiva 2013/37/UE establece políticas de apertura a la información, propiciando la disponibilidad y la reutilización de la información en el sector público con fines privados o comerciales, con restricciones mínimas o nulas de carácter jurídico, técnico y económico, buscando favorecer la circulación de la información no solo para los agentes económicos, sino también entre el público en general.

La nueva Directiva se relaciona también con la Agenda Digital Europea, iniciativa emprendida por la Comisión Europea (2010)² que, a su vez, promueve la comunicación de la Comisión Europea sobre datos abiertos como motor para la innovación, el crecimiento y la gobernanza transparente. Este documento es también un importante referente para la armonización de las normas y prácticas nacionales en materia de reutilización, proporcionando considerables mejoras para la reutilización transfronteriza de los datos³.

Los portales de datos abiertos suponen el elemento más visible de las políticas en materia de *Open Government* y una de las principales herramientas de acceso a los recursos informativos del sector público (Martín, 2011: 113-134). Pueden considerarse, además, como el principal producto tangible en el que se van a ver materializadas las estrategias políticas de las diferentes administraciones públicas en torno al gobierno abierto.

Estos portales se han convertido también en una “ventanilla única” que la administración pone al servicio del ciudadano, evitando la dispersión de contenidos entre las diferentes páginas Web de los organismos públicos. Facilitan, por tanto, el acceso y la reutilización de la información y deberían responder, más allá de ser meras herramientas, repositorios o catálogos de datos, a una vocación de servicio público. En este sentido, hay que indicar que no se trata de generar datos a la carta, sino de poner a disposición de conjuntos de datos que sean mantenidos, que se encuentren actualizados y en formatos estructurados para su consulta y reutilización.

² La Agenda Digital Europea forma parte de la estrategia “Europa 2020” puesta en marcha por la Comisión Europea y promueve entre sus acciones clave la apertura de datos públicos para su reutilización, la simplificación del sistema de licencias para el intercambio de contenidos y la implantación de normas de interoperabilidad. Para más información sobre la Agenda Digital Europea se puede consultar: <http://ec.europa.eu/digital-agenda/>

³ Muchos estados miembros de la Unión Europea han generado sus propias normas, derivadas de la Directiva 2003/98/CE, superando el nivel mínimo fijado en dicha Directiva, pero que han supuesto un obstáculo a la hora de la reutilización de los datos de forma transfronteriza, ya que para poder conocer el alcance del uso de los datos había que acudir a las diversas legislaciones nacionales.

Los reutilizadores de la información defienden el uso de la publicación de datos en crudo o *raw data*. Este tipo de formato está compuesto exclusivamente por datos sin tratar ni catalogar (Pintos; Marín, 2013: 72-77). Para este sector infomediario⁴ resulta de especial interés el conseguir estos ficheros de datos en bruto, pero en formatos que puedan posteriormente gestionar y manipular con la intención de generar productos o servicios de valor añadido.

Como afirma L. Fernando Ramos y otros autores del Grupo Publidoc de la Universidad Complutense de Madrid *“un gran número de estados miembros han iniciado, con mayor o menor entusiasmo, la creación de portales open data con el objetivo principal de hacer accesible a los ciudadanos la información pública disponible y, al mismo tiempo, facilitar a los infomediarios el uso de la información orientada a fines económicos y sociales”* (Ramos y otros, 2012).

2 OBJETIVOS Y METODOLOGÍA

El objetivo general del estudio es ofrecer una valoración sobre la situación actual del portal, identificar buenas prácticas y establecer posibles recomendaciones de mejora. Como objetivos específicos en relación con dicho objetivo general, se formulan los siguientes:

- Presentar una descripción global del portal como proyecto estratégico en el marco de la Agenda Digital para Europa.
- Analizar los conjuntos de datos disponibles a partir de una serie de indicadores relacionados con: editores, formatos disponibles y otros elementos complementarios de análisis.
- Evaluar los formatos de recursos, atendiendo a su condición como materia prima para la reutilización.
- Analizar los diferentes estándares en uso del portal.

Se ha seguido una metodología de investigación de tipo mixto que ha consistido en:

- Un análisis cuantitativo de los conjuntos de datos a partir de los indicadores previamente identificados y seleccionados: conjuntos de datos, editores, editores por tipo de entidad, editores por país,

⁴ Se considera sector infomediario a aquellas organizaciones que basan su modelo de negocio en la reutilización de la información, provenga del sector público o no. Su operativa se basa en el acceso a información que no ha sido producida por ellos y a la que tratan de aportar valor por medio de la creación de productos o servicios que ofrecen a terceros para obtener beneficios.

conjuntos de datos por editor, formatos, formatos reutilizables y no reutilizables. Para poder llevarlo a cabo se ha utilizado el lenguaje de consulta *SPARQL* a través del editor terminal de consultas que ofrece el propio portal⁵.

• Un análisis cualitativo del portal basado en la observación por medio de la navegación del portal, en muchos casos “registro a registro”, la identificación de los elementos principales, los modos de búsqueda y presentación de los resultados, estándares en uso y metadatos disponibles. La evaluación de los formatos de recursos toma como base el esquema de las “5 estrellas de los datos abiertos” (Bernes-Lee, 2006)⁶, que se apoya en un sistema piramidal de asignación de estrellas para determinar el grado de disponibilidad y accesibilidad de los formatos, correspondiendo la asignación de una estrella al grado de menor valor y la asignación de cinco estrellas al de máximo valor (ver tabla 1).

Tabla 1. Esquema de las 5 Estrellas de los datos abiertos.

Fuente: www.w3.org/DesignIssues/LinkedData.html

*	Los datos se encuentran disponibles en la Web. Se pueden consultar y leer, pero no pueden ser utilizados por un software específico para su reutilización. Constituyen por tanto un formato “cerrado” que no es útil para la reutilización. Por ejemplo, los ficheros en formatos PDF.
**	Los datos se encuentran disponibles en la Web en un formato estructurado aunque propietario. Por tanto, es un formato que facilita la reutilización, ya que se pueden procesar, exportar y publicar los datos con facilidad, aun dependiendo del software propietario necesario para su recuperación. Por ejemplo, un fichero en formato XLS (de Excel).
***	Los datos se encuentran en un formato estructurado no propietario. Se trata por tanto de formatos “abiertos” que resultan de gran utilidad para la reutilización. Por ejemplo, un fichero en formato CSV.
****	Utiliza estándares del consorcio W3C para la Web semántica (URI, RDF, SPARQL, etc.) para identificar los objetos de forma inequívoca y puedan ser usados o enlazados desde el exterior.
***** *	Los datos se presentan vinculados a otros datos (de terceros) de manera que se encuentran contextualizados semánticamente.

Este estudio se ha desarrollado en el período comprendido entre enero de 2014 y mayo del 2014.

⁵ Este editor de consultas se encuentra disponible en la zona dedicada a “Datos Vinculados” y concretamente en la dirección: <https://open-data.europa.eu/es/linked-data>

⁶ Este sistema fue introducido por Tim Bernes-Lee, Director del World Wide Web Consortium (W3C) para la evaluación de datos abiertos.

3. RESULTADOS

3.1 EL PORTAL DE DATOS ABIERTOS DE LA UNIÓN EUROPEA: UNA VISIÓN GLOBAL

El portal de datos abiertos de la Unión Europea⁷ puede considerarse como un proyecto estratégico dentro de los objetivos marcados por la Agenda Digital para Europa. Este portal se presenta como un punto de acceso único para una gran diversidad de datos que han sido producidos por las diversas instituciones y organismos de la Unión Europea (ver figura 1).

Figura 1. Captura de la página de inicio del Portal de datos abiertos de la Unión Europea. Consulta: 5 de mayo de 2014



Mediante la Decisión 2011/833/UE de la Comisión, de 12 de diciembre de 2011, relativa a la reutilización de los documentos de la Comisión se insta a lo siguiente:

⁷ El portal se encuentra disponible en la siguiente dirección Web: <https://open-data.europa.eu>

“debe crearse un portal de datos como punto de acceso único a los documentos disponibles para ser reutilizados. Debe adoptarse una disposición que tenga en cuenta la tendencia hacia formatos legibles por máquina. Una importante mejora con respecto a la Decisión 2006/291/CE que consiste en ofrecer en general documentos de la Comisión para su reutilización a través de licencias abiertas de reutilización o sencillas cláusulas de exención de responsabilidad, sin necesidad de efectuar solicitudes individuales” (Unión Europea, 2011).

Con dicha iniciativa se pretendía desarrollar un portal Web que actuara como un único punto de acceso a los conjuntos de datos producidos y mantenidos por los servicios de la Comisión Europea y, por extensión, a los conjuntos de datos producidos y mantenidos por otras instituciones y organismos europeos. El portal de datos de la Unión Europea se abre al público en una versión beta a finales de diciembre del año 2012 y se inaugura oficialmente en enero del año 2013.

La gestión operativa del portal recae en la Oficina de Publicaciones de la Unión Europea y su coordinación en la Dirección General de Redes de Comunicación, Contenido y Tecnologías de la Comisión Europea que es la responsable directa de la política de datos abiertos. Hay que resaltar que dicha política se basa en el principio de que la información pública debe estar a disposición del ciudadano y poder utilizarse y reutilizarse libremente. En este sentido, se promueve el uso de los datos como un medio para fomentar el desarrollo de nuevos proyectos innovadores, aprovechando toda la información semántica que contienen. Tal y como indican en el propio portal Web los “datos se pueden utilizar, reutilizar, enlazar y redistribuir gratuitamente con fines comerciales o no comerciales” (Portal Europeo de Datos Abiertos, 2014).

La arquitectura del portal se basa en soluciones de software libre como el sistema de gestión de contenidos *Drupal* o el sistema para la gestión de datos *CKAN*⁸ y utiliza tecnologías de la Web Semántica para relacionar los datos y presentarlos, como *RDF Virtuoso*⁹ y el editor *SPARQL*¹⁰ para su consulta.

En cuanto a sus contenidos, este recurso es un importante activo de información para el conjunto de estados miembros de la Unión Europea y, por consiguiente, es un importante vehículo de propagación de la información que la Unión Europea genera y pone a disposición de la sociedad en general y de los

⁸ CKAN es un sistema de datos de gran alcance desarrollado por la Open Knowledge Foundation (OKF). CKAN es una solución abierta dirigida a proveedores de datos (gobiernos nacionales, regionales, locales, empresas y organizaciones) que quieran proporcionar datos abiertos y reutilizables.

⁹ RDF Virtuoso es una base de datos de motor híbrido que combina diferentes funcionalidades (servidor de aplicaciones Web, base de datos virtual, servidor de archivos, etc.).

¹⁰ SPARQL (SPARQL Protocol and RDF Query Language) es un lenguaje estandarizado para la consulta de datos RDF de la World Wide Web Consortium (W3C).

investigadores en particular. Con este recurso de acceso fácil y gratuito a los datos, la Comisión Europea también pretende fomentar el uso público de los datos y su reutilización en beneficio de nuevos proyectos innovadores y transfronterizos.

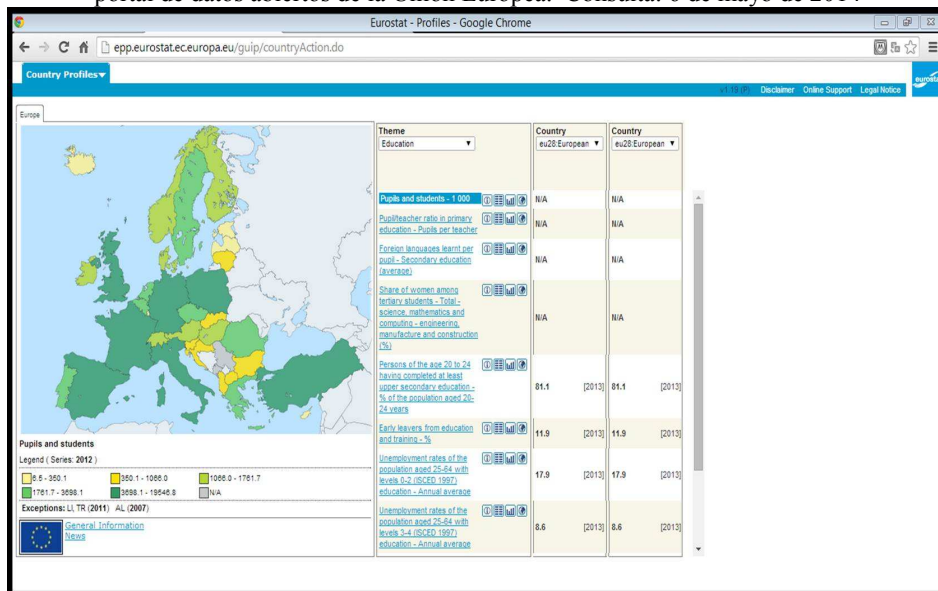
El portal se presenta por medio de un catálogo de metadatos que facilita el acceso a los conjuntos de datos o *datasets* disponibles. Los metadatos son vinculados y normalizados basándose en estándares que facilitan su recuperación y reutilización posterior. Estos conjuntos de datos se encuentran disponibles en formatos que permiten su interpretación humana, siendo también legibles por un ordenador para facilitar su inmediata reutilización. Existen también conjuntos de datos sin procesar que utilizan aplicaciones de visualización que han sido elaboradas con el fin de que puedan ser consultados y reutilizados. Estas aplicaciones han sido creadas por diversos proveedores de datos y conforman una serie de herramientas excepcionales para explorar, interpretar y reutilizar de forma eficiente la información suministrada.

Las aplicaciones se muestran en una galería que incluye 10 herramientas¹¹. Cabe destacar, entre estas aplicaciones, el *Marcador de la Agenda Digital (DG Connect)*, para la creación de gráficos de forma interactiva a partir de la consulta de datos; el *Visor de Mapas del Centro Europeo de Información Forestal (EFDAC)*, una aplicación Web cartográfica personalizada que permite al usuario visualizar, navegar y consultar mapas y otros conjuntos de datos geográficos; el *Sistema Europeo de Información sobre Incendios Forestales (EFFIS)* que sirve de apoyo a los servicios de protección contra incendios de la Unión Europea y ofrece información actualizada y fiable sobre los fuegos que se producen en Europa; el *EuroStat LänderProfile* que incluye perfiles de los países a partir de un amplio grupo de indicadores de interés sobre cada país: población, educación, seguridad, condiciones de vida, economía, medio ambiente, etc., ofreciendo además tablas, gráficos y mapas interactivos (ver figura 2); el *Atlas Europeo del Mar* que facilita el aprendizaje sobre los mares y las costas del viejo continente o la aplicación *CubeViz*, la primera en estar disponible desde la versión beta del portal, y se trata de un navegador facetado para datos estadísticos que emplea el vocabulario *RDF Data Cube*¹², el más avanzado para representar datos estadísticos en RDF.

¹¹ A fecha de consulta de 6 de mayo de 2014 (Fuente: Portal de datos abiertos de la Unión Europea).

¹² RDF Data Cube es una recomendación de la W3C que permite relacionar datos multidimensionales, como las estadísticas en la Web, de tal manera que permite vincular conjuntos de datos relacionados y conceptos. Para más información se puede consultar: www.w3.org/TR/2014/REC-vocab-data-cube-20140116/

Figura 2. Captura de una consulta en la aplicación *EuroStat LänderProfile* disponible en el portal de datos abiertos de la Unión Europea. Consulta: 6 de mayo de 2014



Los datos que se facilitan a través del portal pueden utilizarse y reutilizarse con fines comerciales o no comerciales, buscando impulsar su potencial económico. Su reutilización, de forma gratuita, conlleva únicamente la citación de la fuente. Toda la información que se ofrece está sujeta a una cláusula de exención de responsabilidad, a un aviso de derechos de autor y a las normas aplicables de protección de datos personales¹³.

Con relación a la interfaz del portal, podemos observar que se encuentra disponible en 22 idiomas. Incluye una sencilla navegación por medio de menús y un formulario de búsqueda simple que permite filtrar los resultados a partir de la inclusión de todas las palabras de la búsqueda, de alguna de las palabras o de la frase exacta.

El portal también se constituye como una fuente de importante valor para la comparación de datos entre los diversos organismos productores y fomenta el establecimiento de relaciones entre los mismos y la comunidad de usuarios. Además, como herramienta, dispone de un área exclusiva para proveedores de datos.

¹³ Para más información sobre este aviso legal puede consultarse la siguiente dirección: http://ec.europa.eu/geninfo/legal_notices_es.htm

3.2 ESTADO DE LOS CONJUNTOS DE DATOS

El total de conjuntos de datos disponibles declarado en la página principal del portal es de 6.597¹⁴. No obstante, al realizar una revisión de este dato, observamos que en realidad se refiere a registros o entradas del repositorio y no a conjuntos de datos como recursos reales disponibles. Dichas entradas o registros deberían tener asociado al menos un conjunto de datos como recurso en el campo correspondiente, pero en algunos casos no es así (ver figura 3).

Figura 3. Captura de un registro de la *Dirección General de Educación y Cultura*. En dicho registro se puede observar como anomalía que no existe ningún recurso disponible.
Consulta: 12 de mayo de 2014

The screenshot shows the 'Portal de datos abiertos de la Unión Europea' interface. The main heading is 'Eve publishes results of projects in the fields of Education, Training, Culture, Youth and Citizenship'. The 'Recursos' section is highlighted with a red box and contains the text '(ninguno)'. Other sections include 'Licencia: Europa Legal Notice', 'Registro en el catálogo' (with dates 2012-12-12 and 2013-12-17), and 'Proponga un conjunto de datos' with a link to request data. The footer indicates the last update was on 17/10/2013.

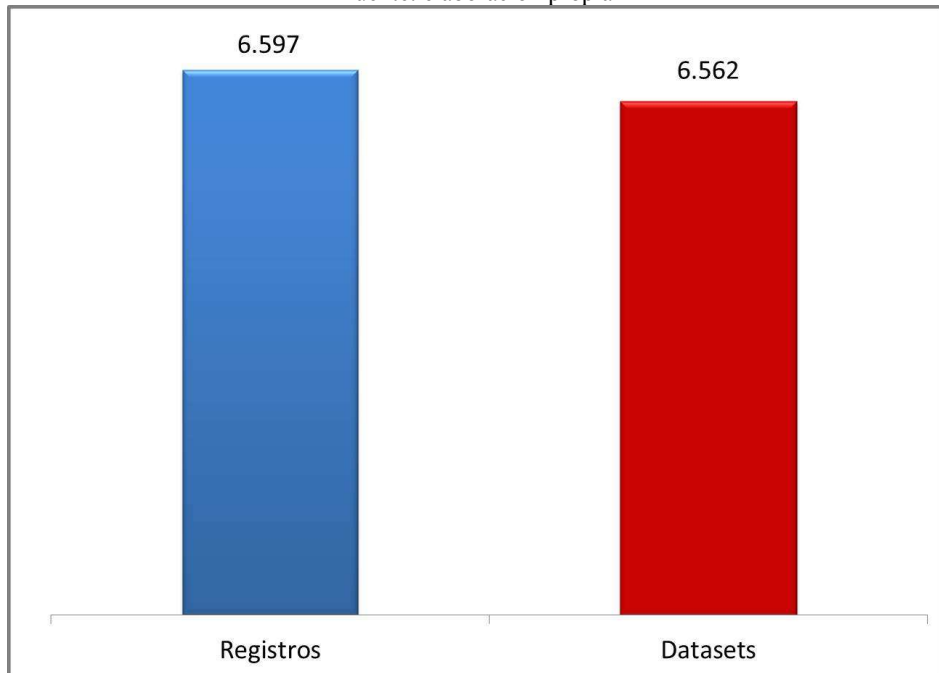
El dato obtenido sobre los conjuntos de datos vinculados a un recurso real disponible, independientemente del formato o formatos, es de 6.562 (ver gráfico 1).

En este sentido, cabe destacar el caso de la *Agencia Europea de Valores y Mercados* cuyo registro es de 27 conjuntos de datos y ofrece solamente 4 recursos

¹⁴ A fecha de consulta de 12 de mayo de 2014 (Fuente: Portal de datos abiertos de la Unión Europea).

enlazados o el caso de la *Dirección General de Educación y Cultura* que tiene un registro de 3 conjuntos de datos y no figura ningún recurso enlazado (ver figura 3).

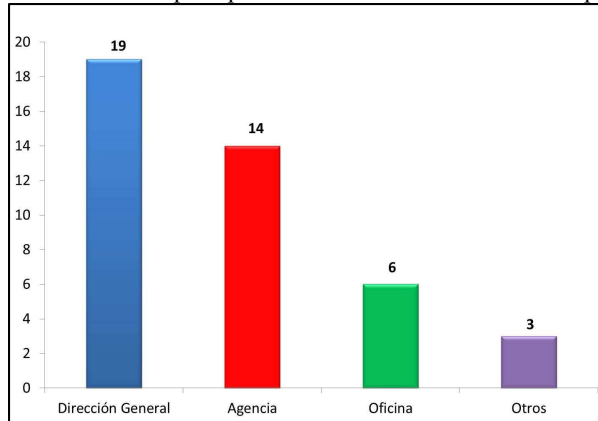
Gráfico 1. Comparativa entre número de registros disponibles y conjuntos de datos disponibles.
Fuente: elaboración propia



3.3 EDITORES

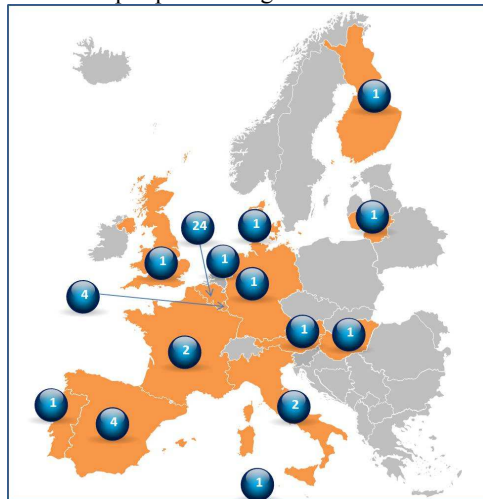
En cuanto a editores nos encontramos con un número total de 41 editores. La mayor representación, teniendo en cuenta el tipo de entidad, es el de las *Direcciones Generales*, seguido de las *Agencias*¹⁵ y *Oficinas* (ver gráfico 2).

¹⁵ Las Agencias de la Unión Europea son entidades jurídicas independientes de las instituciones de la Unión Europea creadas para llevar a cabo tareas específicas bajo la normativa de la UE.

Gráfico 2. Editores por tipo de entidad. Fuente: elaboración propia

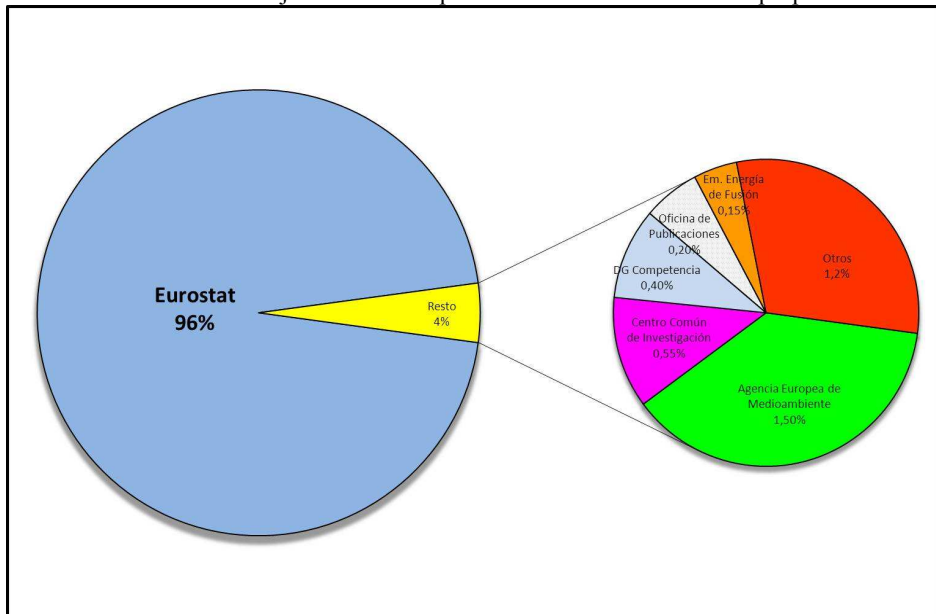
Si nos centramos en la función de estos organismos podemos observar que los de mayor representación son aquellos relacionados con economía y finanzas (7), seguidos de los que se ocupan de asuntos sociales (5), empresas (5), asuntos institucionales (5), ciencia y tecnología (4), asuntos marítimos y pesca (3).

En cuanto al país donde se encuentran ubicados estos organismos podemos ver que los más representados, al ser la mayoría *Direcciones Generales*, se encuentran en Bélgica y, más concretamente, en Bruselas, ya que muchas de estas direcciones son dependientes de la Comisión Europea. En cuanto al país de origen de las Agencias: Luxemburgo (4), España (4), Francia (2) e Italia (2), son los países más representados, aunque existe representación de hasta 15 países (ver figura 4).

Figura 4. Editores por país de origen. Fuente: elaboración propia

Hay que mencionar que de las más de 40 Agencias vinculadas a la Unión Europea solamente 14 de ellas aportan conjuntos de datos a este portal. Uno de los datos más significativos es el número de conjuntos de datos que aporta cada editor. En este caso destaca la *Oficina de Estadística de la Unión Europea (Eurostat)*¹⁶ como principal editora de conjuntos de datos, con un 96% de los mismos; destaca también la aportación de la Agencia Europea de Medio Ambiente (ver gráfico 3).

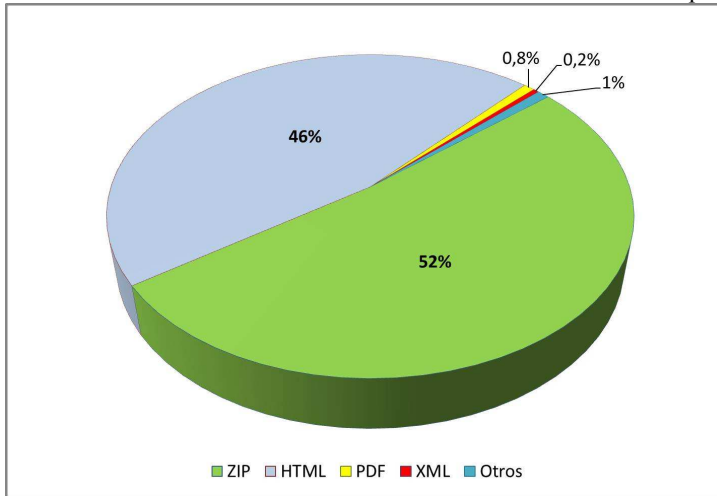
Gráfico 3. Conjuntos de datos por editor. Fuente: elaboración propia



3.4 FORMATOS

En cuanto a los formatos de distribución de estos recursos se encuentran en su gran mayoría en más de un formato, siendo el formato comprimido .zip el más habitual, seguido del formato .pdf, .xml, .xls y otros formatos (ver gráfico 4).

¹⁶ Eurostat produce datos estadísticos sobre la Unión Europea. Su misión es proporcionar a la Unión Europea las estadísticas que permitan comparaciones entre los países y regiones. Para más información sobre Eurostat: <http://epp.eurostat.ec.europa.eu/>

Gráfico 4. Relación de formatos de distribución. Fuente: elaboración propia

Teniendo en cuenta que los conjuntos de datos pueden ser distribuidos en más de un formato, buscando precisamente la posible reutilización de los datos, nos encontramos con un total de 6.571 ficheros en diferentes formatos, donde se descartan los 5.688 registros en formatos .html, debido a que estos recursos son realmente enlaces a las páginas Web de inicio de los editores y, en muy pocos casos, llevan a zonas de descarga.

Sobre la base de esta panorámica inicial, hemos de señalar que los datos que ofrece el propio portal en el apartado de formatos de recursos¹⁷, no reflejan con precisión la presencia en dicho portal de formatos que favorecen la reutilización. En este sentido, un análisis inicial de estos datos, muestra un alto porcentaje (47%) de documentos en formatos .html y .pdf, (formatos de distribución que no son los más adecuados para la reutilización) que contrasta, *a priori*, con el bajo porcentaje mostrado de recursos en formatos estructurados o delimitados: xml, rdf, xls, csv (1,2% de los recursos).

Un desglose de los formatos integrados en los archivos comprimidos .zip (que representan el 52% de los datos de distribución), desvela una presencia superior de formatos estructurados y delimitados que los proporcionados *a priori* en el apartado mencionado. Para poder conocer el formato del contenido de estos .zip es preciso explorar cada recurso y comprobar su formato específico, en algunos casos incluso se requiere la descarga del recurso, lo que resulta poco ágil y eficiente. No obstante, hemos comprobado que el 99% de estos ficheros comprimidos proviene

¹⁷ <https://open-data.europa.eu/es/data/dataset>

de *Eurostat* y contienen ficheros con formato *sdmx-ml*¹⁸. Estos formatos *sdmx-ml* son formatos basados en xml, por lo tanto, formatos estructurados y reutilizables. En consecuencia, se observa una presencia mucho más alta de formatos estructurados que la que se muestra en la información previa.

3.5 EVALUACIÓN DE LOS FORMATOS MEDIANTE LA ADAPTACIÓN DEL ESQUEMA DE LAS “5 ESTRELLAS”

En el Portal de datos objeto de estudio, todos los recursos presentan una URI, además los datos se encuentran vinculados entre ellos por palabras clave, con independencia de quien sea el editor; desde esta perspectiva, el portal en su conjunto se situaría en el nivel de las 5 estrellas.

Sin embargo, cuando se analizan individualmente los recursos, se constata que no todos cumplen con esta condición. Como ya se ha comentado, existen conjuntos de datos o recursos que se presentan en varios formatos. Por ejemplo, un recurso se puede presentar como un documento pdf y, además, como un fichero .xls. En estas circunstancias, ese conjunto de datos cumple, según el esquema citado, con los requisitos propios del nivel de 2 estrellas. Solamente a aquellos conjuntos de datos que contienen como único formato el formato .pdf, se les ha otorgado 1 estrella. La mayoría de los formatos pueden considerarse formatos estructurados no propietarios (3 estrellas). También existen formatos rdf, owl y sparql (4 estrellas). Los formatos estructurados no propietarios que se acompañan, además de otros formatos tipo rdf, han sido clasificados con 4 estrellas (tabla 2).

Tabla 2. Evaluación de formatos según adaptación del esquema de 5 estrellas.

Fuente: elaboración propia.

*	1,1%
**	0,4%
***	98%
****	0,5%

En esta evaluación vuelve a ser determinante el peso de los contenidos de *Eurostat*, clasificados con 3 estrellas. Aunque debemos indicar, como aspecto negativo, que existe una muy alta relación de editores con formatos evaluados con 1 estrella o, dicho de otra manera, del total de editores, el 41% los que publican únicamente en formato .pdf (1 estrella).

¹⁸ SDMX se corresponde con una iniciativa internacional para intercambio de datos y metadatos estadísticos.

3.6 ESTÁNDARES

El portal se presenta como un entorno de datos vinculados donde se toman como base los principios de *Linked data*¹⁹ y otras técnicas de la Web Semántica. Los datos vinculados son un método normalizado para representar gran variedad de temas, facilitando su relación y su uso. Resultan también útiles para el desarrollo de nuevas aplicaciones que permiten combinar datos de origen heterogéneo.

Además, el portal se apoya en diferentes estándares que permiten su adecuada descripción y una eficaz recuperación, como por ejemplo el lenguaje de consulta *SPARQL* para RDF. *SPARQL* es una recomendación de la W3C que define la forma de expresar consultas a través de diversas fuentes de datos, siempre que los datos se almacenen de forma nativa en formato RDF. Por otro lado, también se apoya en el *Data Catalog Vocabulary (DCAT)*, otra recomendación de la W3C, que conforma un vocabulario en formato RDF diseñado para facilitar la interoperabilidad entre los diferentes catálogos de datos publicados en la Web.

Tomando como base *DCAT* se ha creado un vocabulario de metadatos que puede consultarse por medio de una hoja de datos o por medio de una ontología en formato *OWL*²⁰. Este vocabulario se ha generado con el objetivo de que sea compatible con *Asset Description Metadata Schema (ADMS)*. *ADMS* es un esquema de *DCAT* usado para describir activos semánticos. Estos activos semánticos o, simplemente “activos”, facilitan aquellos componentes altamente reutilizables para este tipo de proyectos u otros basados en metadatos (como por ejemplo, esquemas XML, modelos de datos genéricos, etc.) y también para recuperar los datos de referencia (como por ejemplo, listas de códigos, taxonomías, diccionarios, vocabularios, etc.) utilizados para el desarrollo y la descripción de los valores en los sistemas de *eGovernment* o Gobierno abierto.

3.7. OTROS ELEMENTOS COMPLEMENTARIOS DE ANÁLISIS

Con relación al formulario de búsqueda podemos observar que no podemos utilizar un tesauro para la consulta y el filtrado de resultados, como podría ser Eurovoc²¹, que posibilite la localización de los recursos con independencia del

¹⁹ *Linked Data* forma parte de la familia de estándares de la Web Semántica desarrollados por el World Wide Web Consortium (W3C). Este término hace referencia a un conjunto de buenas prácticas para la publicación e interconexión de datos estructurados en la Web, introducidas por Tim Bernes-Lee y que han llegado a ser conocidas como los *principios de linked data*. Para más información sobre estos principios puede consultarse la siguiente dirección: <http://www.w3.org/DesignIssues/LinkedData.html>

²⁰ OWL es el acrónimo de Web Ontology Language, un lenguaje de marcado que se utiliza para publicar y compartir datos utilizando ontologías en la Web.

²¹ EuroVoc es un tesauro multilingüe y multidisciplinar que abarca la terminología de los ámbitos de actividad de la Unión Europea, con especial hincapié en las labores parlamentarias.

idioma en el que se encuentran. En este sentido, hemos de señalar que en las búsquedas realizadas en el contexto del análisis, se obtuvieron resultados diferentes según se planteara la consulta con términos en español o con términos en inglés. Sirva como ejemplo para ilustrar dicha consideración que, en uno de los supuestos planteados, cuando se introdujo el término en español, “agua” no se obtuvieron resultados y, sin embargo, al introducir el término en inglés “water”, se obtuvieron 109 resultados. En relación con lo expuesto, al disponer de la interfaz en español, incluido el formulario de búsqueda, la consulta por parte de los usuarios no expertos o no habituales del portal suele ser también en el mismo idioma, lo puede generar como resultados continuos silencios documentales y, en consecuencia, frustración en el usuario que desconoce esta situación.

Para finalizar el análisis de elementos complementarios, hemos de señalar que este portal también presenta un *top ten* de los conjuntos de datos más vistos. Nos ha parecido relevante completar dicha relación con información relativa a los editores y formatos (ver tabla III). Destacamos que en el primer puesto se encuentra la “Memoria de Traducción” de la Dirección General de Traducción, un conjunto de datos que incluye traducciones en 23 idiomas del acervo comunitario (legislación europea, tratados, reglamentos, directivas, etc.).

Tabla 3. Top 10 de los recursos más vistos desde el lanzamiento del Portal.
Fuente: Portal Europeo de datos abiertos, 12 de mayo de 2014 (elaboración propia)

Top ten: lo más visto			
Titulo	Editor	Consultas	Formatos
Memoria de traducción	DG de Traducción	4.648	tmx-xml
Mapa de altitudes de Europa	Agencia Europea de Medio Ambiente	2.090	binary data
Longitud total de líneas ferroviarias	Eurostat	1.708	sdmx-ml
Tesoro Eurovoc	Oficina de Publicaciones	1.513	rdf+xml
Generación de residuos por sectores	Eurostat	1.263	sdmx-ml
Comercio cruzado de mercancías por carretera	Eurostat	1.245	sdmx-ml
Porcentaje de personas empleadas con habilidades de usuario de las TIC	Eurostat	1.223	sdmx-ml
Servicios de telecomunicaciones: acceso a redes	Eurostat	911	sdmx-ml
Camas para cuidados psiquiátricos por hospitales	Eurostat	829	sdmx-ml
CORDIS - proyectos de investigación de la UE en el programa FP7 (2007-2013)	Oficina de Publicaciones	765	xls csv

EuroVoc está disponible en 23 lenguas oficiales de la Unión Europea. Para más información: <http://eurovoc.europa.eu/>

4 CONCLUSIONES Y RECOMENDACIONES

El Portal de datos abiertos analizado constituye en su concepción, sin lugar a dudas, un recurso muy valioso como punto de acceso único a datos elaborados por instituciones y organismos de la Unión Europea. Como tal, se configura como una iniciativa que propicia la disponibilidad de información del sector público y su reutilización, con fines comerciales o no comerciales, en el contexto del movimiento de *datos públicos abiertos*. En consecuencia, dado que la finalidad primordial a la que responde es el fomento del uso público de los datos y su reutilización, se perfila como una plataforma para el surgimiento de nuevos proyectos de carácter innovador y transfronterizo.

En línea con el análisis realizado de dicho activo de información, consideramos que sería recomendable mejorar o definir de forma más precisa algunas de sus características para contribuir a reforzar mejor su finalidad como incentivo para la reutilización. En este sentido, sería muy positivo que los proveedores de datos indicaran con mayor precisión el formato de los conjuntos de datos que suministran, aunque estos vayan comprimidos. Conocer este dato es fundamental para definir y formular proyectos para la reutilización de la información.

Aunque existen diversas opiniones sobre cuál es el mejor formato para la reutilización, por lo general y, siempre que sea posible, lo más eficiente suele ser la publicación en diversos formatos y que, al menos uno de ellos, sea estructurado abierto (csv, xml, etc.). El coste de este proceso para el productor suele ser muy pequeño pero, desde el punto de vista del usuario final, puede representar una gran diferencia.

También consideramos que resultaría muy positiva la incorporación de un tesoro multilingüe que facilitaría la búsqueda. En relación con esa mejora de las características, sería muy útil disponer en la navegación del portal de una organización de los contenidos por grandes áreas temáticas.

Finalmente, como activo de información vivo, aún en fase de desarrollo y en constante transformación, confiamos en que, progresivamente, se irá produciendo un aumento de la participación de nuevos proveedores de datos que, sin lugar a dudas, redundará en un incremento cuantitativo del número de conjuntos de datos y aplicaciones y también en una mejora cualitativa basada, desde nuestro punto de vista, en una mayor diversificación de editores y contenidos.

5 BIBLIOGRAFÍA Y FUENTES

- BERNES-LEE, Tim. Linked data. Is your Linked Open Data 5 Star? <<http://www.w3.org/DesignIssues/LinkedData.html/>> [Consulta: 5/5/2014]
- COMISIÓN EUROPEA (1998). “La información del sector público: un recurso clave para Europa. Libro Verde sobre la información del Sector Público en la Sociedad de la Información”. COM(1998)585 [en línea] <ftp://ftp.cordis.europa.eu/pub/econtent/docs/gp_es.pdf> [Consulta: 5/05/2014]
- COMISIÓN EUROPEA (2011). Comunicación desde la Comisión al Parlamento Europeo, el Consejo, el Comité Económico y Social y el Comité de las Regiones. Datos abiertos. Un motor para la innovación, crecimiento y gobernanza transparente. Bruselas, 12 de diciembre de 2011. COM(2011)882 final.
- KROES, Neelie (2009). “Public data for all – opening up Europe’s public sector”. [en línea]. <http://ec.europa.eu/commission_2010-2014/kroes/en/blog/public-data-for-all-%E2%80%93-opening-up-europes-public-sector> [Consulta: 5/05/2014]
- MARTIN, Yolanda (2011). “Política europea de reutilización de la información del sector público. De la norma jurídica al portal de datos abiertos”, en *Revista Universitaria Europea*, nº 19, diciembre 2013, pp. 113-134.
- OBAMA, B. (2009). “Transparency and open government. Memorandum for the heads of executive departments and agencies”, en *Federal Register*, nº 15, vol. 74, pp. 4685-4686.
- PINTOS, Carlos; MARÍN, José Luis (2013). Open data. Reutilización de la información pública. Madrid: INAP; Netbiblio, pp. 72-77.
- Portal Europeo de datos abiertos*. <<https://open-data.europa.eu/es/data/>> [Consulta: 7/5/2014]
- RAMOS SIMÓN, L. Fernando, ARQUERO AVILÉS, Rosario, BOTEZAN, Iuliana, COBO SERRANO, Silvia, SALA JIMÉNEZ, Andrea, SANCHEZ JIMÉNEZ, Rodrigo, VALLE GASTAMINZA, Félix del (2012). “De la reutilización de la información del sector público a los portales de datos abiertos en Europa”, en *Bid: textos universitarios de biblioteconomía i documentació*, diciembre.
- UNIÓN EUROPEA. Directiva 2003/98/CE del Parlamento Europeo y del Consejo de 17 de noviembre de 2003 relativa a la reutilización de la información del sector público. *Diario Oficial de la Unión Europea L 345*, 31 de diciembre de 2003, pp. 90-96.
- UNIÓN EUROPEA. Directiva 2013/37/UE del Parlamento Europeo y del Consejo, de 26 de junio de 2013, por la que se modifica la Directiva 2003/98/CE relativa a la reutilización de la información del sector público. *Diario Oficial de la Unión Europea L 175*, 27 de junio de 2013, pp. 1-8.

UNIÓN EUROPEA. Decisión 2011/833/UE de la Comisión, de 12 de diciembre de 2011, relativa a la reutilización de los documentos de la Comisión. *Diario Oficial de la Unión Europea L 330*, 14 de diciembre de 2011, pp. 39-42.

Los autores pertenecen al Grupo de investigación Publidoc-UCM, dirigido por el Dr. L. Fernando Ramos Simón y la Dra. Rosario Arquero Avilés. El artículo es un resultado de investigación englobado en el proyecto “Organización del acceso, uso y reutilización de la información del sector público. Hacia la consolidación de una industria de la información”, financiado por el Ministerio de Economía y Competitividad (referencia: CSO2010-17451).