

El portal de datos abiertos de la Unión Europea en el contexto del *Open Data Government*

ROSARIO ARQUERO AVILÉS*
GONZALO MARCO CUENCA
Universidad Complutense de Madrid, España

INTRODUCCIÓN

Los movimientos alrededor de los datos abiertos en el sector público, que se producen desde el año 2009 en Europa, han fomentado la creación de múltiples portales en el territorio de la Unión Europea.¹ El portal de datos abiertos de la Unión Europea constituye un punto de acceso único para una gran diversidad de datos que han sido producidos por instituciones y otros organismos de la Unión Europea. Los datos abiertos representan una excelente solución para la difusión de la información y su reutilización posterior, facilitando el acceso al público en general y a los investigadores.

El portal promueve el uso de los datos como un medio para fomentar el desarrollo de nuevos proyectos innovado-

* Agradecimiento a la Dirección General de Asuntos del Personal Académico (DGAPA, UNAM), proyecto “Buenas prácticas en el acceso a la información gubernamental” (PAPIIT IN 403113)

res 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.”²

En este contexto, el portal de datos abiertos de la Unión Europea se enlaza con la Directiva 2013/37/UE, del Parlamento Europeo y del Consejo, relativa a la reutilización de la información en el sector público.³ La Directiva, que modifica y actualiza una norma anterior, pretende establecer políticas de apertura a la información que propicien la disponibilidad y la reutilización generalizadas de la información en el sector público con fines privados o comerciales, con restricciones mínimas o nulas de carácter jurídico, técnico y económico, y que deben favorecer la circulación de la información no sólo 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) 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.⁴ La Directiva 2013/37/UE también supone un esfuerzo importante para conseguir una armonización mínima de las normas y prácticas nacionales en materia de reutilización, y proporciona una considerable mejora para la reutilización transfronteriza de los datos.

Hay que tener en cuenta que autorizar la reutilización de los datos de los organismos públicos confiere a los mismos un valor añadido para los reutilizadores, para los usuarios finales y para la sociedad en general; además, contribuye a la generación y desarrollo de nuevos conocimientos. De igual modo, supone para los organismos públicos una mejora asociada a la propia transparencia del organismo e implica un

aumento de la responsabilidad sobre los datos que favorece la mejora de la calidad de la información que se difunde.

CONTENIDOS DEL PORTAL

El portal de datos de la Unión Europea 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 pública de la sociedad en general y de los investigadores. Con este recurso de acceso fácil y gratuito a los datos, la Comisión Europea pretende fomentar el uso público de los datos y su reutilización en beneficio de nuevos proyectos innovadores y transfronterizos.

Los datos que se facilitan a través del portal pueden utilizarse y reutilizarse con fines comerciales o no comerciales, con lo cual se busca impulsar su potencial económico. Su reutilización, de forma gratuita, conlleva únicamente la citación de la fuente (véase *Figura 1*). 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 personales.

Figura 1

Captura de la página de inicio del portal de datos abiertos de la Unión Europea



De forma general, como ya se ha comentado, los datos se pueden reutilizar de manera gratuita y solamente un pequeño conjunto de excepciones incluyen determinadas condiciones para su reutilización, la mayoría relacionadas con la propiedad intelectual de terceras partes.

El portal, disponible en 22 idiomas, constituye también una fuente de importante valor para la comparación de datos entre los diversos organismos productores y fomenta el

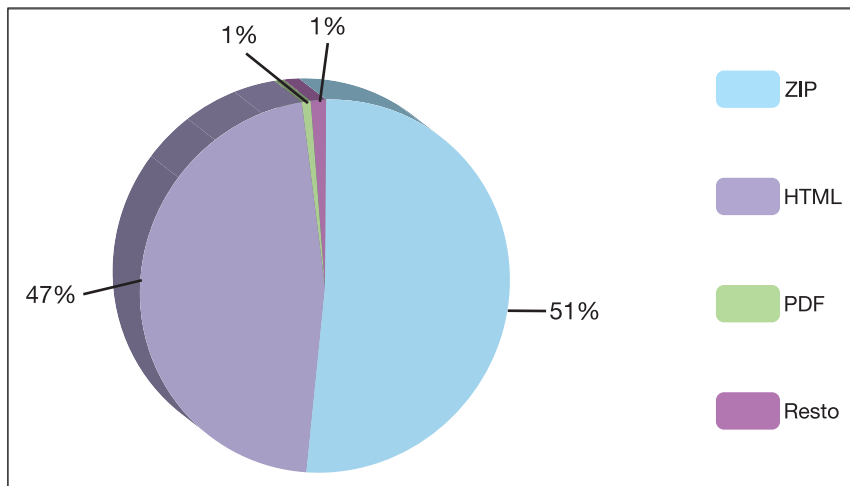
establecimiento de relaciones entre éstos y la comunidad de usuarios. Además, como herramienta, dispone de un área exclusiva para proveedores de datos.

En cuanto a los datos, el portal presenta un catálogo de metadatos que facilita su acceso. 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, y son también legibles por un ordenador para facilitar su inmediata reutilización.

El total de recursos o conjuntos de datos disponibles es de 6.565.⁵ En cuanto a los formatos de distribución de estos recursos, se encuentran en su gran mayoría en más de un formato. El formato comprimido .zip es el más habitual, seguido del formato .html y otros, tal y como se muestra en la *Figura 2*.

Figura 2

Formatos de distribución de los recursos del portal de datos abiertos Europeo



Fuente: Elaboración propia, a partir del portal de datos abiertos de la Unión Europea

Partiendo de esta panorámica inicial, hemos de señalar que los datos que ofrece el propio portal, en el apartado de formato de recursos, no reflejan con precisión la presencia de formatos que favorecen la reutilización.⁶

En este sentido, un análisis inicial de estos datos muestra un alto porcentaje (48%) de documentos en formatos .html y .pdf (formatos de distribución que no son los más adecuados para la reutilización), lo cual contrasta, *a priori*, con el bajo porcentaje mostrado de recursos en formatos estructurados o delimitados: xml, rdf, xls, csv, txt (1% de los recursos).

Además, si se tiene en cuenta que los archivos comprimidos .zip (que representan el 51% de los datos de distribución en el apartado de formato de recursos mencionado) pueden considerarse un contenedor de otros formatos o ficheros, se confirma que la información preliminar de formatos de recursos que ofrece el propio portal no reflejan la representación real de los formatos en el portal.

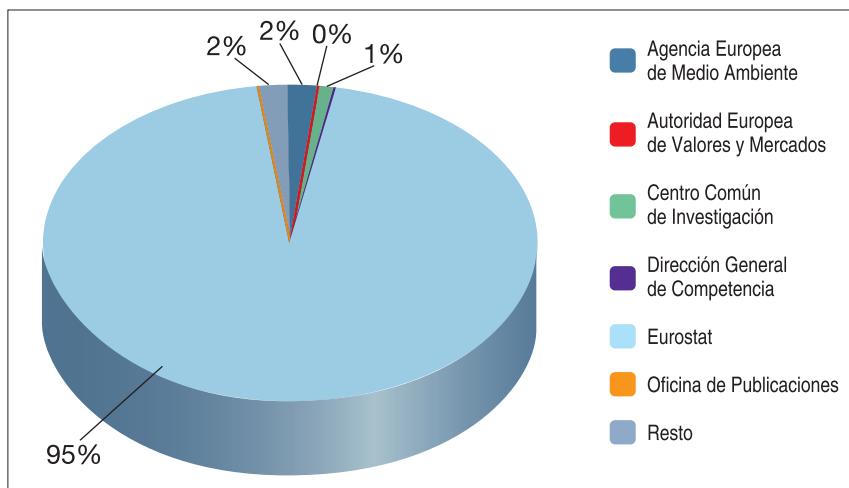
Un desglose de los formatos integrados en los archivos comprimidos .zip desvela una presencia superior de formatos estructurados y delimitados que los proporcionados *a priori* en el apartado mencionado.

De este modo, para conocer el formato del contenido de estos .zip es preciso explorar cada recurso para comprobar su formato específico, lo que resulta poco ágil cuando tenemos que utilizar diferentes recursos. Con esta operación, y teniendo en cuenta que la mayoría de los datos proviene de la *Oficina de Estadística de la Unión Europea (Eurostat)*, podemos observar que se da una presencia mucho más alta de formatos estructurados y delimitados que la que se muestra en la información previa. En consecuencia, pensamos que el detalle en el desglose de los formatos resulta realmente importante en un contexto en el que se busca

el fomento de la reutilización. En esta línea, resulta crucial concienciar a los proveedores para que faciliten los metadatos lo más completos posibles y, en especial, a aquellos relacionados con los formatos.

En cuanto a los editores de los datos, *Eurostat* es la principal editora de conjuntos de datos, con un 95% de los éstos. Al respecto, destaca también la aportación de la *Agencia Europea de Medio Ambiente* (ver *Figura 3*). En total, participan más de 40 editores diferentes que presentan conjuntos especializados de todo tipo.

Figura 3
Editores de conjuntos de datos del portal de datos abiertos Europeo



Fuente: Elaboración propia, a partir del portal de datos abiertos de la Unión Europea

La información que se puede encontrar para describir cada conjunto de datos, los cuales conforman el catálogo de metadatos anteriormente mencionado, está constituida por los siguientes elementos: *editor*, *descripción del recurso*, *recursos* (con enlaces para la descarga e información sobre el formato en el que se pueden encontrar los datos, también

se incluye una etiqueta que representa el formato del fichero que los contiene, que suele ser un fichero comprimido .zip o una página .html), su *Uniform Resource Identifier (URI)*, su estado, un identificador, un tipo de conjunto de datos, la fecha de modificación, la cobertura temporal del recurso, los datos de contacto del editor, información sobre la licencia, palabras clave, e información sobre su registro en el catálogo.

DATOS VINCULADOS: ESTÁNDARES

Este portal se presenta también como un entorno de datos vinculados donde se toman como base los principios de *Linked data* y otras técnicas de la Web Semántica. Los datos vinculados son un método normalizado de representar gran variedad de temas, lo cual facilita 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, si los datos se almacenan de forma nativa en formato RDF.⁷ 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.⁸

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*.

Este vocabulario se ha generado con el objetivo de que sea compatible con *Asset Description Metadata Schema (ADMS)*. *ADMS* es un perfil 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, etcétera) y también para recuperar los datos de referencia (como, por ejemplo, listas de códigos, taxonomías, diccionarios, vocabularios, etcétera) utilizados para el desarrollo y la descripción de los valores en los sistemas de *e-government* o Gobierno abierto.⁹

APLICACIONES DE VISUALIZACIÓN

Existen conjuntos de datos sin procesar que utilizan aplicaciones que han sido elaboradas en torno a los datos, para 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 e incluyen 10 herramientas.¹⁰

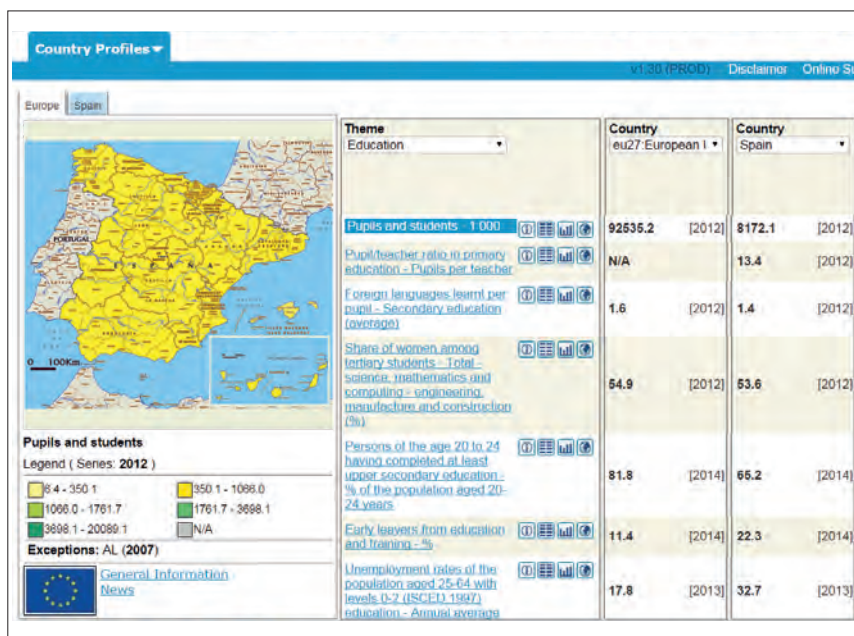
Cabe destacar, entre otras aplicaciones, el *Marcador de la Agenda Digital (DG Connect)*, capaz de crear gráficos de forma interactiva a partir de la consulta de datos, el *Visor de Mapas del Centro Europeo de Información Forestal (EFDAC)*, una aplicación web cartográfica personalizada que permite al usuario visualizar, navegar y consultar mapas y otros conjuntos de datos geográficos, el *EuroStat LänderProfile*, que incluye perfiles de los países a partir de un amplio grupo de indicadores de interés sobre cada país: población,

La información gubernamental y el acceso a la información...

educación, seguridad, condiciones de vida, economía, medio ambiente, etcétera, ofreciendo además tablas, gráficos y mapas interactivos (ver *Figura 4*) o la aplicación *CubeViz*, un navegador facetado para datos estadísticos que emplea el vocabulario *RDF Data Cube*, el más avanzado para representar datos estadísticos en RDF.

Figura 4

Captura de una consulta en la aplicación *EuroStat LänderProfile* disponible en el portal de datos abiertos de la Unión Europea



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 como usuarios de dicho activo de información y teniendo en cuenta, tal y como se ha señalado, que dicho portal se encuentra en fase de desarrollo, pensamos que sería recomendable mejorar o definir de forma más precisa algunas de sus características para contribuir a reforzar mejor su finalidad en el incentivo a la reutilización. En este sentido, pensamos que sería muy positivo mejorar el desglose en la presentación de formatos de recursos disponibles, dado que la presencia real de formatos reutilizables es superior a la mostrada en el apartado de formato de recursos del propio portal.

También consideramos que resultaría muy positiva la incorporación de ayudas sobre cómo usar el buscador o información complementaria sobre las características idiomáticas del portal. En esta línea hemos de señalar que, aunque el formulario de consulta es multilingüe, no es posible formular consultas con palabras clave en español.

En línea con esa mejora de las características, sería muy útil en la navegación del portal disponer 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 en una mayor diversificación de editores y contenidos.

REFERENCIAS BIBLIOGRÁFICAS

1. L. F. Ramón Simón *et al.* (2012), “De la reutilización de información del sector público a los portales de datos abiertos en Europa”, en *bid: textos universitaris de biblioteconomia i documentació*, 29 [en línea], <http://bid.ub.edu/29/ramos2.htm>. DSGS
2. Unión Europea (2013), Portal de datos abiertos de la Unión Europea [apartado: “¿De qué trata este Portal de datos?”], [en línea], <https://open-data.europa.eu>
3. Parlamento Europeo (2013), “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,” en *Diario Oficial de la Unión Europea* (27/06/2013) [en línea], <http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2013:175:0001:0008:ES:PDF>
4. Comisión Europea (2011), “Comunicación de la comisión al parlamento europeo, al consejo, al comité económico y social europeo y al comité de las regiones. Datos abiertos: un motor para la innovación, el crecimiento y la gobernanza transparente” COM(2011) 882 final [en línea], <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:ES:PDF>
5. Unión Europea (2013), *Op. cit.*
6. *Ibíd.*
7. W3C Recommendation. SPARQL Query Language for RDF [en

línea], <http://www.w3.org/TR/rdf-sparql-query/>

8. W3C Recommendation. Data Catalog Vocabulary (DCAT) [en línea], <http://www.w3.org/TR/vocab-dcat/>
9. W3C Working Group Note. Asset Description Metadata Schema (ADMS) [en línea], <http://www.w3.org/TR/vocab-adms/>
10. Unión Europea (2013), *Op. cit.*