

ARTÍCULO DE REVISIÓN

Los datos abiertos: definición técnica de un concepto clave para la Transformación Digital

*Open Data: a Technical Definition of a Key Concept
for the Digital Transformation*

Antonio F. Rodríguez Pascual

afrodriguez@fomento.es • <https://orcid.org/0000-0002-0111-1349>

Celia Sevilla Sánchez

cssanchez@fomento.es • <https://orcid.org/0000-0002-0548-9277>

Juan Manuel Rodríguez Borreguero

juanm.rodriguez@cnig.es • <https://orcid.org/0000-0003-1254-581X>

CENTRO NACIONAL DE INFORMACIÓN GEOGRÁFICA, ESPAÑA

Recibido: 2020-05-28 • Aceptado: 2020-06-16

RESUMEN

Los datos se han convertido en el recurso más valioso, al ser la materia prima en la que se basan los negocios legales más lucrativos, especialmente los datos sobre la actividad de las personas, su traza completa en la web y en la superficie de la Tierra. Qué hacemos en la red en cada momento y desde dónde.

El acceso a ese *big geodata* se encuentra actualmente custodiado por unas pocas empresas y para hacer posible una transformación digital de las organizaciones en la que haya igualdad de oportunidades y un mercado no enraizado, sería necesario que esa galaxia de datos se publicase en forma de datos abiertos, a menudo datos geográficos abiertos.

Un gran número de organizaciones e iniciativas recomiendan en muchos ámbitos la publicación de datos abiertos. Sin embargo, no hay una definición clara y precisa, universalmente aceptada de qué son exactamente datos abiertos, ni un conjunto de requisitos técnicos que permita tener una lista de chequeo para verificar y en su caso, certificar, que un conjunto de datos se está publicando como datos abiertos.

En este artículo, se proporciona una definición conceptual y teórica de datos abiertos y se resume el contenido de la norma española UNE 148004:2019 que propone una definición verificable de datos geográficos abiertos, extrapo-

lable a otros tipos de datos. Finalmente, se plantean algunas líneas futuras de trabajo y se esbozan unas conclusiones.

PALABRAS CLAVE: datos abiertos; interoperabilidad; licencias de uso; política de datos; transformación digital.

ABSTRACT

Data has become the most valuable resource, being the raw material on which the most lucrative legal businesses are based, especially data about the activity of people, their tracks on the web and on the surface of the Earth. What we do on the web at any times and from any place.

Access to that Big GeoData is currently controlled by a few companies and to make possible a digital transformation of organizations in which there is equality of opportunities and an un-rarefied market, it is necessary that this galaxy of data were published in the form of open data, often open geographic data.

A large number of organizations and initiatives recommend the publication of open data in many initiatives. However, there is no universally accepted clear and precise definition of what open data exactly is, nor a set of technical requirements in order to have a checklist to verify and, if necessary, to certify that a dataset is being published as open data.

This article provides a conceptual and theoretical definition of open data and summarizes the content of the Spanish standard UNE 148004: 2019 that proposes a verifiable definition of open geographic data, extrapolated to other types of data. Finally some future lines of work are included and a set of conclusions are outlined.

KEYWORDS: *data policy; digital transformation; interoperability, use licences; open data.*

«Pero tan pronto como hube adquirido algunas nociones generales de la física... creí que conservarlas ocultas era grandísimo pecado, que infringía la ley que nos obliga a procurar el bien general de todos los hombres».

René Descartes (Discurso del método, 1637)

INTRODUCCIÓN

Parece que fue el matemático y emprendedor británico Clive Humby quien acuñó en el 2006 la frase “Los datos son el nuevo petróleo” (Arthur, 2013), cinco palabras muy citadas, una de cuyas

elaboraciones más completas se publicó en *The Economist* el 6 de mayo de 2017 (*The Economist*, 2017) bajo el título *The world's most valuable resource is no longer oil, but data* (El recurso más valioso del mundo ya no es el petróleo, sino los datos), argumentando que en la economía actual, los datos juegan el papel que antes jugaba el petróleo en la economía vigésima y predigital: el de la materia prima cuyo flujo genera los negocios legales más lucrativos. El artículo continuaba exponiendo que las cinco mayores empresas tecnológicas, *Amazon, Apple, Facebook, Google y Microsoft*, cuyos beneficios anuales superaban en ese momento los 100 000 millones de dólares, gestionan ingentes cantidades de datos sobre nosotros, sin apenas control ni supervisión. Saben lo que compramos, buscamos en la web, compartimos con otros, recomendamos y preferimos, lo que unido a las transacciones de nuestra tarjeta de crédito y nuestra posición geográfica en cada momento, fácilmente extraíble de nuestro móvil, puede permitir que nuestro comportamiento sea estadísticamente predecible, con los beneficios que ello supone.

La falta de control sobre las actividades de esas grandes empresas es posible que afecte, de un modo que aún desconocemos, no solo a nuestra privacidad, que ya nunca volverá a ser lo que era, sino también a la innovación, los comportamientos sociales y al libre mercado. En esta última posibilidad se basaba el texto de *The Economist* para sostener la necesidad de nuevas leyes antimonopolio para garantizar la igualdad de oportunidades.

Por otro lado, puesto que se empieza a considerar que los datos de nuestra posición son datos personales, tal y como demostró *The New York Times* (Valentino-De Vries, 2018) y reconoce el Reglamento General de Protección de Datos (Unión Europea, 2016), al definir datos personales como “información sobre una persona física identificada o identificable”, probablemente sería necesario disponer de nuestra autorización para explotar esos datos.

Por otro lado, de poco sirve que pidamos transparencia a nuestros gobiernos si nos olvidamos de pedirla también a las empresas que gestionan información sobre nosotros cuya manipulación tiene consecuencias de largo alcance y parece que un mínimo exigible sería estar informados de qué explotación se hace de nuestros datos, qué negocio se genera y quién gestiona esa información.

Todos los aspectos mencionados apuntan en la misma dirección: probablemente lo más adecuado sería publicar esos datos como datos abiertos.

Otra notable fuente de macrodatos es la constituida por los gobiernos, que a través de sus departamentos gestionan ingentes volúmenes de información sobre nosotros y los más variados campos de actividad, que por su gran utilidad se demanda que se compartan como datos abiertos (Unión Europea, 2003, 2013 y 2019). Baste señalar que el portal de datos abiertos de Reino Unido¹, abierto en 2010 y una de las iniciativas pioneras en ese campo, publica en mayo de 2020 nada menos que algo más de 54 000 conjuntos de datos generados por autoridades públicas en el ejercicio de sus funciones y el portal europeo de datos abiertos² tiene en la misma fecha un catálogo de más de un millón de conjuntos de datos de la misma naturaleza.

¹ [https:// data.gov.uk](https://data.gov.uk)

² <https://www.europeandataportal.eu/>

Por otro lado, no hay que olvidar como fuentes de datos de gran valor a las universidades y empresas del sector privado, que entre ambas constituyen actualmente en el portal de datos abiertos de España un 79,4 % (21 459 de un total de 27 037 conjuntos de datos) así como a los proyectos colaborativos o de *crowdsourcing*, en los que de nuevo, la información geográfica juega un papel muy relevante gracias a la iniciativa *OpenStreetMap*, que se está consolidando como uno de los recursos cartográficos más usados a nivel global en todo tipo de proyectos⁴.

Por último, varios autores (Singleton, *et al*, 2016) defienden los datos abiertos como pilar básico de las Ciencias de información geográfica, que para poder alcanzar la reproductibilidad de experimentos necesitan basarse en aplicaciones de fuentes abiertas y datos abiertos.

Vemos pues que los seis tipos de fuentes de datos mencionados, los gigantes tecnológicos, los gobiernos, la universidad, el sector privado en general, los proyectos colaborativos y los proyectos de *GI Science*, todas ellas de gran relevancia, constituyen fuentes potenciales de datos abiertos atendiendo a la demanda existente.

Los datos abiertos han sido reconocidos por varios autores (Carrara, *et al*, 2015) (Mention, 2019), como uno de los recursos clave y una palanca decisiva para que tanto empresas privadas como públicas completen el proceso de su transformación digital.

Vamos a analizar a continuación algunos problemas que arrastra la conceptualización precisa y rigurosa de qué son exactamente los datos abiertos.

UNA SITUACIÓN CONTRADICTORIA

Se da el hecho curioso de que mientras un considerable número de iniciativas y organizaciones recomiendan, especialmente a las autoridades públicas y a través de informes, declaraciones e iniciativas, la publicación de datos abiertos, el tanto por ciento de conjuntos de datos publicados en condiciones abiertas es todavía sorprendentemente bajo.

Efectivamente, como hitos más importantes en el impulso de los datos abiertos podemos mencionar:

- El Convenio de Aarhus (UNECE, 1998) sobre el derecho de los ciudadanos a acceder a los datos ambientales.
- El informe PIRA (PIRA, 2000) que llegaba a cuantificar el incremento de beneficio económico diferencial que suponía la publicación abierta de datos gubernamentales.
- La *Budapest Open Access Initiative* (*Open Society Institute*, 2001) sobre el acceso abierto a los artículos de investigación.
- El esquema de cinco estrellas para datos abiertos (Berners-Lee, 2006), que incluye la generación de datos enlazados (*linked data*) en su concepción de datos abiertos.
- La subsecuente fundación del *Open Data Institute*⁵ en 2012 para la promoción y difusión de los datos abiertos.

³ <https://open.data.es>

⁴ <https://welcome.openstreetmap.org/about-osm-community/consumers/>

⁵ <https://theodi.org/>

- La iniciativa *World Bank Open Data* del Banco Mundial, que incluye un programa de ayudas económicas para que los países publiquen datos abiertos (Herzog, 2017).
- La *Open Data Charter (UK Cabinet Office, 2013)* firmada por los países del G8 en el 2013, que incluía el compromiso de publicar un plan de actuación en el 2013 y elaborar un informe anual.
- La *International Open Data Charter (ODC, 2015)*, apoyada por el G20 y Naciones Unidas en el 2015, que amplió la anterior, firmada actualmente, en mayo del 2020, por 22 gobiernos nacionales y 51 gobiernos locales.
- Las ya mencionadas tres Directivas Europeas sobre Reutilización de la Información del Sector Público (Unión Europea, 2003, 2013 y 2019).
- El *European Interoperability Framework v2 (Comisión Europea, 2017)*, que recomienda publicar datos abiertos, a menos que haya un impedimento legal como protección de datos personales, confidencialidad o derechos de autor de terceros.
- La Declaración de Dubai (Naciones Unidas, 2018), que apoya decididamente la publicación como datos abiertos de los datos relacionados con los Objetivos de Desarrollo Sostenible y la Agenda 2030.

Sin embargo y a pesar de la sucesión de recomendaciones y declaraciones de interés en favor de los datos abiertos, se observa que a pesar de que existe una tendencia generalizada de publicación de datos abiertos, el tanto por ciento de conjuntos de datos abiertos publicados es todavía notoriamente bajo:

- El *Open Data Barometer (W3C, 2017)* y el *Global Open Data Index (Open Knowledge Foundation, 2017)* sitúan el tanto por ciento de datos públicos que se publican en todo el mundo como datos abiertos en cifras que no sobrepasan el 11 %.
- Un estudio presentado por el *Joint Research Center* en la Conferencia *Inspire 2018* (Hernández, *et al*, 2018) concluía que de 29 000 registros de metadatos de datos y servicios que contenía el catálogo *Inspire* en el año 2016, solo el 25 % incluía la declaración de una licencia de uso abierta.
- En España, en un estudio realizado en 2018 por el Subgrupo de Trabajo de la IDEE (Rodríguez, 2018) sobre políticas de datos y licencias de uso, sobre un total de 90 productores de datos geográficos oficiales integrado en la IDE de España, se encontró que solo el 28 % de las organizaciones publicaban esos datos bajo una licencia abierta.

METODOLOGÍA

Asumimos la hipótesis de que una de las causas que frenan la expansión de los datos abiertos puede ser, precisamente, que no existe una definición clara, precisa y universalmente admitida de qué son datos abiertos. Esa falta de una definición técnica permite que no sea fácil determinar si una organización está realmente publicando datos abiertos o no, dificulta la confección y publicación de indicadores de datos abiertos y hace que en ocasiones sus responsables no sepan cómo mejorar en la apertura de datos. Efectivamente creemos que

la frase, dudosamente atribuida a Lord Kelvin⁶, 'Lo que no se define no se puede medir. Lo que no se mide, no se puede mejorar' es tan popular en el campo de la ingeniería porque encierra una gran verdad.

En las próximas secciones, recopilaremos las principales definiciones de datos abiertos en uso, para tratar de mostrar la variedad existente y avanzar hacia una definición precisa y operativa de ese concepto. Expondremos las ideas principales en las que se basa la norma española de datos geográficos abiertos (UNE, 2019), fácilmente extrapolable a otros dominios temáticos y junto a otras conceptualizaciones que parece necesario tomar en cuenta, trazaremos las líneas maestras para llegar a una definición definitiva y formularemos un conjunto de recomendaciones y conclusiones.

Como se verá más adelante, se han tenido en cuenta las iniciativas de generación de indicadores mundiales de datos abiertos más significativas y las iniciativas de datos abiertos más relevantes en la actualidad, también en el ámbito mundial.

DESARROLLO

DEFINICIONES DE DATOS ABIERTOS

Si bien se encuentra cierta similitud en cómo se entiende el concepto de datos abiertos, creemos que hay cierta falta de unanimidad en algunos de los matices que se incluyen en las definiciones que se manejan.

Portal Europeo de Datos

El *European Data Portal*⁷ concibe los datos abiertos como datos a los que cualquiera puede acceder, utilizar de cualquier manera que desee, transformar y compartir con otros, incluso de manera comercial. Aspectos como el formato y la legibilidad automática considera que deben ser tenidos en cuenta, pero que hacen los datos más usables, no más abiertos.

Queda la duda de si “de cualquier manera” incluye, por ejemplo, el atribuirse la autoría de los datos, lo que no suele permitirse, y excluye la obligación de usar formatos abiertos y la legibilidad por máquina, que sí suele incluirse en la idea de datos abiertos.

International Open Data Charter

Menciona que son datos digitales disponibles con las características técnicas y legales para que puedan ser libremente utilizados, reutilizados y redistribuidos por cualquiera, en cualquier momento y lugar (ODC, 2015).

Introduce el carácter digital de los datos; habría que definir el término “libremente” con precisión y añade el matiz de la universalidad (en cualquier momento y lugar).

⁶ <https://zapatopi.net/kelvin/quotes/>

⁷ <https://www.europeandataportal.eu/elearning/en/#/id/co-01>

European Interoperability Framework v2

Se refiere a la idea de datos que deben estar disponibles libremente para su uso y reutilización, salvo que estén afectados por restricciones, por ejemplo, debidas a datos personales, confidencialidad o derechos de propiedad intelectual (Comisión Europea, 2017).

Con lo que menciona también un término tan borroso como “libremente” y algunas excepciones razonables para no publicar unos datos como datos abiertos, no para dejar de publicarlos y seguir declarándolos como abiertos.

Directivas RISP

Las tres directivas europeas sobre Reutilización de la Información del Sector Público no definen explícitamente qué datos son datos abiertos (Unión Europea, 2003, 2013 y 2019).

Open Knowledge Foundation

Proporciona la definición técnica más precisa y concreta (*Open Knowledge Foundation*, 2016), que incluye una lista de requisitos técnicos detallados y bien descritos sobre las licencias a utilizar, la disponibilidad de la obra como un todo, sin coste alguno, el empleo de formatos abiertos y legibles por máquina, una larga lista de usos permitidos, la condición de no discriminar a ningún grupo de usuarios y algunas restricciones admisibles, esencialmente reconocimiento de la autoría y compartir igual. Define todos los conceptos que utiliza y ha conseguido si no un consenso unánime, sí la adhesión de iniciativas muy importantes. En ella se basan el *Global Open Data Index* (*Open Knowledge Foundation*, 2017), el *Open Data Institute* y el *Open Data Handbook*⁸.

A pesar de su concreción y detalle, adolece de un conjunto de pruebas abstractas (*abstract test suite*), es decir una descripción detallada del método de prueba a aplicar para verificar de manera objetiva si un conjunto de datos se está publicando como datos abiertos y que pueda servir de base para poder llegar a certificarlo.

Es decir, se presenta el problema de definir un indicador de apertura de datos, tarea en la que sería necesario establecer con rigor y de manera nítida tres aspectos: qué se va a determinar (la apertura de datos), qué indicador o indicadores se definen y el método de determinación. Tres puntos necesarios en todo proceso de medida o determinación cuantitativa, de acuerdo a lo que especifican, entre otras normas, las UNE 66175:2003 (UNE, 2003) e ISO 19157:2013 (ISO, 2013).

Una definición teórica que coincide con la experiencia

Una de las formulaciones que nos parece mejor articuladas alrededor de la idea de qué datos abiertos son los datos que pueden usarse libremente, es la que nos dio Barbara Ubaldi, *Head of Digital Government and Open Data Unit* de la OCDE, durante la presentación *Informe del Sector Infomediario 2018* de ASIEDIE (Asociación Multisectorial de la Información⁹) cuando

⁸ <https://opendatahandbook.org/>

⁹ <http://www.asedie.es/asedie.html>

le preguntamos “¿qué son exactamente los datos abiertos?”: *datos que se publican sin barreras económicas, técnicas ni legales para su uso y reutilización.*

Definición que parece oponerse radicalmente a la idea de datos cerrados (datos que no es posible utilizar debido a barreras que lo impiden) y que coincide con nuestra experiencia de publicación de datos geospaciales en el Centro Nacional de Información Geográfica (CNIG), que depende del Instituto Geográfico Nacional (IGN) de España.

Después de siete años de publicar datos con una política basada en no permitir usos comerciales y definida en la Orden Ministerial FOM/956/2008 (Ministerio de Fomento, 2008), se implementaron en el Centro de Descargas del CNIG tres medidas para la publicación de datos geospaciales del IGN:

- En mayo de 2015 se habilitó la descarga múltiple de tantos ficheros como deseara el usuario con un único clic.
- También en mayo de 2015 se suprimió la necesidad de que el usuario se registrase con una dirección de correo electrónico y una clave como requisito para poder descargar datos.
- En diciembre de 2015, se puso en práctica una nueva política de datos, basada en una licencia compatible con una CC BY 4.0, que permite usos comerciales (Ministerio de Fomento, 2015).

Contrariamente a lo que esperábamos, las dos primeras medidas tuvieron un impacto mucho mayor que la tercera tanto en el número de ficheros descargados como en el volumen total de las descargas (véanse la figura 1 y la tabla 1).

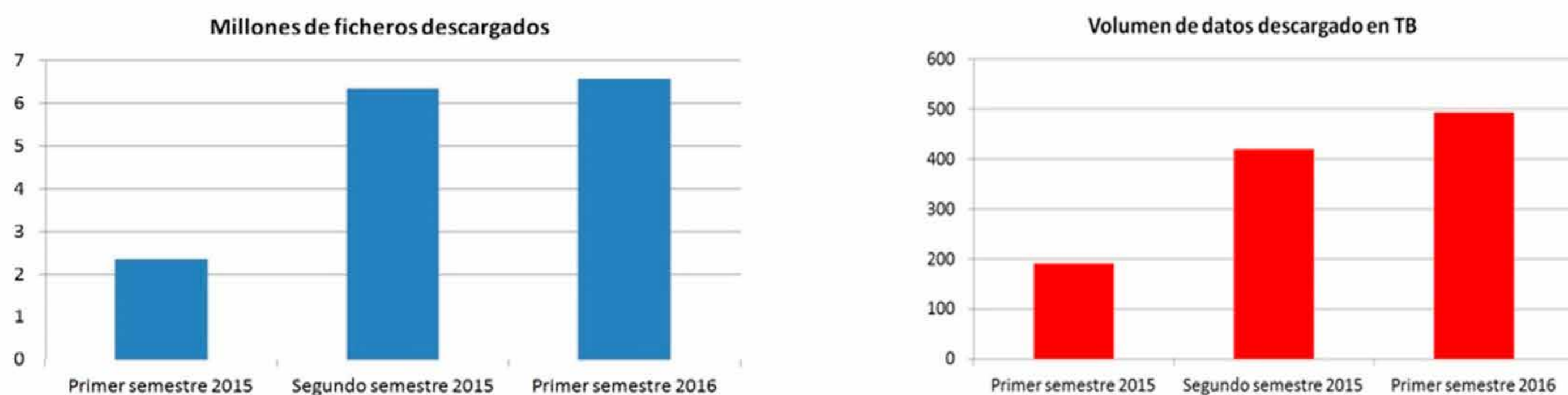


Figura 1. Número de ficheros y volumen de datos descargados en el Centro de Descargas del CNIG en los semestres primero y último de 2015 y primero de 2016.

Tabla 1. Cifras totales de terabytes y ficheros descargados en el Centro de Descargas del CNIG en los semestres primero y último de 2015 y primero de 2016

	1er. semestre de 2015	2do. semestre de 2015	1er. semestre de 2016
Millones de ficheros descargados	2,36	6,36	6,57
TB de datos descargados	190,87	420,29	493,72

Por lo tanto, efectivamente, nuestra experiencia fue que la supresión de dos barreras, que pueden parecer a primera vista poco relevantes, tuvo un impacto considerablemente mayor

que un cambio significativo en los usos permitidos por la licencia de datos utilizada. Cualquier barrera, por pequeña que sea, puede tener un impacto significativo en la difusión de los datos. Por tanto, si se pretende publicar datos abiertos, con el objetivo de que unos datos lleguen al mayor número posible de usuarios y campos de aplicación, vale la pena realizar un esfuerzo en eliminar todo tipo de barreras para la circulación de esos datos.

Nuestra evolución en la publicación de datos puede decirse que ha consistido en que, hace años, en una primera aproximación, pensamos ingenuamente que publicar datos abiertos significaba fundamentalmente ofrecer datos gratuitos; luego aprendimos que la licencia, es decir, los usos permitidos y las condiciones a cumplir para utilizar los datos son la esencia de la apertura de datos, y ahora, nos hemos dado cuenta de que la idea esencial es eliminar de manera radical todo tipo de barreras.

Hacia una definición operativa: la UNE 148004

Sin embargo, si bien la definición anterior es conceptualmente irreprochable, parece que define una situación utópica, un ideal hacia el que avanzar, ya que cualquier pequeño inconveniente puede ser percibido como una barrera por el usuario y plantea algunas cuestiones de orden práctico como ¿hasta qué punto minimizar? ¿cómo cuantificar y medir esa minimización?

Por otro lado, seguíamos sin disponer de una definición práctica que incluyese un *Abstract Test Suite*, un conjunto de pruebas objetivas y concretas que permitiesen la verificación y certificación de un conjunto de datos como datos abiertos.

Esos fueron los dos motivos, junto a la inexistencia de una norma ISO o de cualquier otro organismo de normalización, que llevaron al Comité Técnico de Normalización 148 de UNE, titulado *Información geográfica digital*, a elaborar una norma española de datos geográficos abiertos durante el año 2017, que finalmente se aprobó a principios del 2018 como la Norma UNE 148004:2018 *Datos geográficos abiertos* (UNE, 2018).

Esa norma está basada y es deudora de la citada mencionada definición de datos abiertos establecida (*Open Knowledge Foundation*, 2016), el *European Interoperability Framework v2* (Comisión Europea, 2017) y el Esquema Nacional de Interoperabilidad (RD 4/2010, 2010) de España.

Como paso previo para sintetizar el planteamiento de la norma vamos a definir con precisión algunos conceptos que resultan clave para comprender el contenido del texto normativo:

- **Estándar:** solución técnica que disfruta de una posición dominante en un sector de actividad determinado.

Efectivamente, el estándar viene determinado sencillamente por la moda en un campo de actividad determinado, lo que se usa, lo que de hecho se utiliza en la práctica y está tan extendido que es necesario tenerlo en cuenta si se pretende definir una norma que atienda a las necesidades de los usuarios. No es necesario que sea la solución más utilizada, sino que tenga una comunidad de usuarios suficientemente grande, estable y sólida como para asegurar su permanencia. Esos conceptos conllevan cierto grado de subjetividad y, efectivamente, no siempre es fácil decidir qué es un estándar y qué no. Hay estándares definidos por organizaciones de estandarización, como el consorcio W3C y

el OGC, otros por empresas privadas, como el formato *shapefile* o el ECW, y otros sencillamente establecidos por el uso, como la práctica de nombrar los mapas mediante el tema, la extensión y la escala (por ejemplo *Mapa forestal de España 1: 200 000*).

- **Estándar abierto de nivel 1:** estándar cuya descripción es pública y gratuita o tiene un coste que no supone una dificultad de acceso, cuyo uso y aplicación no está condicionado al pago de un derecho de propiedad intelectual o industrial y para el que existe al menos un *software* libre que lo lee, véase (Real Decreto, 2010). En este primer nivel de apertura se suprimen las barreras más limitantes para el uso de un estándar, el que tanto su descripción como su aplicación sean gratuitas y el que haya aplicaciones que lo procesen. Se permite que la descripción esté disponible en una publicación de pago, si es tan barata como para que su coste no suponga una barrera, lo que supone introducir cierto grado de subjetividad en el concepto.

No hay que confundir estándar con norma, solución técnica elaborada y aprobada por un organismo oficial de normalización, como ISO en el ámbito internacional, CEN en el europeo, UNE en el español, AFNOR en el francés, DIN en el alemán, etc.

- **Estándar abierto de nivel 2:** estándar abierto de nivel 1 que además cumple la condición de haber sido aprobado y ser mantenido por una organización con fines de interés general, de manera que su evolución está gestionada sobre la base de procedimientos abiertos de toma de decisiones en los que pueden participar todas las partes interesadas y se consensuan las decisiones por mayoría, véase (Comisión Europea, 2017). Al imponer esa condición, se pretende asegurar que la evolución futura del estándar atienda a los intereses de la comunidad y no a los intereses particulares de una empresa u organización en particular, intereses legítimos y respetables, que pueden coincidir o no con los de la comunidad de actores implicados.

La norma española UNE 148004:2018 *Datos geográficos abiertos* establece una definición normalizada, precisa y clara de qué se entiende por datos geográficos abiertos y proporciona un conjunto de requisitos para publicar un conjunto de datos geográficos como datos geográficos abiertos.

Se basa en algunos principios generales relativos a la publicación de datos abiertos:

- El Principio de eliminación radical de todo tipo de barreras para el uso y reutilización de los datos. Es una idea muy general, basada en la definición utópica de Ubaldi antes mencionada.
- El Principio de no discriminación de ningún grupo de usuarios, campo de aplicación, uso pretendido y sector de actividad. El dar facilidades a algún tipo de uso, por ejemplo, a la investigación, rompe ese principio porque supondría una barrera o dificultad para el resto.
- El Principio de neutralidad tecnológica, que se basa en la idea de no discriminar, directa o indirectamente, a ningún usuario en función de la opción tecnológica que haya elegido (sistema operativo, navegador, dispositivo, idioma, entorno cultural...). En realidad, es un caso particular del principio anterior.

Estos principios se mantienen como directrices generales a seguir en la implementación de la apertura de datos, en la que se distinguen cuatro estadios diferentes (datos geográficos disponibles, documentados, bajo una licencia abierta y en un formato abierto) y dos niveles de apertura, basados en los dos niveles de estándar abierto definidos.

1. **Datos geográficos disponibles:** son datos geográficos que se pueden descargar desde una dirección web, mediante un protocolo de comunicaciones estándar y abierto, y en formato digital, descargables sin pérdidas respecto de la información original y de una sola vez, de un solo golpe o como mucho, en fragmentos tan grandes como permita la tecnología disponible en el momento. Deben ser datos de descarga gratuita y se permite cubrir los costes marginales de preparación y copia en soporte digital, si existe esa posibilidad y se realiza a petición del usuario.
2. **Datos geográficos documentados:** son datos que están acompañados de los ítems descriptivos mínimos para que sea posible su interpretación correcta y transformación en la información codificada en ellos sin errores ni ambigüedades. Esa descripción puede proporcionarse en forma de metadatos, preferiblemente conforme a (ISO, 2014), como especificaciones de producto de datos, preferiblemente conforme a (ISO, 2007), directamente en la página de descarga, en un fichero de documentación también descargable o de cualquier otra manera que resulte cómoda y evidente.

Los ítems descriptivos que se consideran necesarios son:

- El idioma o idiomas de los datos.
 - El juego de caracteres en el que están codificados los datos.
 - El formato o formatos en los que están disponibles, junto con su descripción, que puede darse de manera explícita o mediante una URL.
 - La fecha de los datos, que ha de ser la de la última actualización o la fecha de generación de los datos, de manera que en cualquier caso describa cuándo los datos representan un modelo de una parte del mundo real.
 - El Sistema de Referencia por Coordenadas (SRC).
3. **Datos geográficos bajo una licencia abierta:** se define como licencia abierta la que cumple las siguientes condiciones:
 - 3.1 Que sea una licencia estándar o tipo, es decir que está definida de manera pública en Internet, de modo que cualquier usuario pueda acceder a su descripción y características, y cualquier proveedor de datos puede sindicarse a ella invocando la correspondiente URL.
 - 3.2 Que sea de aceptación implícita, es decir que no sea necesario aceptarla explícitamente y la mera descarga de los datos implique la aceptación por el usuario de sus condiciones.
 - 3.3 Que permita usos comerciales, la generación de obras derivadas, copia, modificación, compilación, separación en partes y difusión de los datos originales, de cualquiera de sus partes y de las obras derivadas.
 - 3.4 Que no suponga ningún tipo de discriminación por grupo de usuarios, sector de aplicación, uso pretendido, solución tecnológica o cualquier otra circunstancia o motivo.

3.5 Que sea no revocable y universal, para que el usuario tenga seguridad jurídica de que las condiciones no van a cambiar en el futuro ni se van a ver alteradas en función del país, territorio o área geográfica en la que se utilicen los datos.

Se considera que la licencia sigue siendo abierta aunque exija alguna o todas de las condiciones opcionales siguientes:

1. El reconocimiento de la autoría original, siempre que no resulte oneroso, es decir demasiado difícil y costoso, para lo cual lo mejor es incluir la fórmula que se elija para ese reconocimiento en la misma licencia.
 2. El que los datos se compartan igual, es decir bajo la misma licencia, si es que se vuelven a difundir, tanto los datos originales como otros posibles datos derivados.
 3. El que las obras derivadas que se generen se difundan con una licencia abierta, aunque no sea exactamente la misma.
 4. El que se distingan los datos derivados de los originales, con expresiones del tipo «obra derivada de», «datos derivados de» o similares.
4. En un formato abierto, para que un formato sea abierto se exige: en primer lugar, que sea un formato digital, no se admiten documentos analógicos; en segundo lugar, que sea procesable de manera automática para poder realizar la explotación pretendida para esos datos, y en tercer lugar, que sea un estándar abierto.

La explotación pretendida de unos datos depende de su formato y de su modelo. Por ejemplo, un fichero de datos ráster puede ser de tal naturaleza que los procesos habituales que pueden realizarse sobre él no requieran intervención humana y en ese caso, sería procesable automáticamente. Sin embargo, un conjunto de datos vectoriales que se publica en forma de imagen rasterizada no sería procesable automáticamente de manera directa, porque no se podrían ejecutar las funciones de análisis y tratamiento habituales sobre datos vectoriales.

En cuanto a qué se entiende por un formato estándar abierto, se aplican las definiciones dadas anteriormente y se contemplan dos niveles:

- a. Formato abierto de nivel 1, el que cumple las condiciones anteriores y además es un estándar de nivel 1, es decir es un estándar cuya descripción es pública y gratuita o tiene un coste que no supone una dificultad de acceso, cuyo uso y aplicación no está condicionado al pago de un derecho de propiedad intelectual o industrial y existe al menos un *software* libre que lo lee. Por ejemplo los formatos *shapefile* y *ECW*.
- b. Formato abierto de nivel 2, el que cumple las condiciones anteriores, es un formato abierto de nivel 1 y además, es un estándar de nivel 2, es decir cumple la condición de haber sido aprobado y ser mantenido por una organización con fines de interés general, de manera que su evolución está gestionada sobre la base de procedimientos abiertos de toma de decisiones en los que pueden participar todas las partes interesadas y se consensuan las decisiones por mayoría. Por ejemplo los formatos *KML*, *GML*, *GeoJSON* y *Geopackage*.

Líneas de trabajo e iniciativas pendientes

Algunas de las líneas de trabajo e iniciativas pendientes que se desprende como consecuencia de lo expuesto son las siguientes:

- Documentación adecuada y rigurosa de las condiciones de uso de datos y servicios en los metadatos de datos y servicios, en las especificaciones de productos de datos y en las autodescripciones (*Capabilities* y equivalentes) de los servicios web, a cargo de los productores datos y proveedores de servicio.
- Establecimiento de formatos de datos geográficos eficientes, eficaces, usables y avanzados en el seno de las organizaciones de normalización y estandarización, ante el uso mayoritario de formatos definidos por entidades privadas, tal y como se ha encontrado en (SGT Política de Datos, 2019), probablemente porque generan mayor satisfacción en los usuarios.
- Apoyo legislativo explícito de los gobiernos a los datos abiertos, una vez que parece existir un amplio consenso sobre los beneficios que se derivan para la sociedad en su conjunto (véase la sección «Una situación contradictoria» de este artículo) y prácticamente no existen voces discrepantes.
- Definición de una norma similar a la UNE 148004, preferiblemente internacional, que defina con precisión y rigor, incluyendo un Abstract Test Suite, qué es exactamente un servicio abierto. Existe también una definición de servicio abierto (Open Knowledge Foundation) que constituye una buena aproximación al problema, pero que no se ha generado con el grado de consenso que se producen las normas. Es un paso necesario en coherencia con el cambio de paradigma que supone pasar de una arquitectura orientada a datos a una arquitectura orientada a servicios.
- Mantenimiento de las principales iniciativas de generación de indicadores anuales de datos abiertos, como *Global Open Data Index* (Open Knowledge Foundation, 2017), que lamentablemente interrumpió su serie temporal en 2016 y no describe todos los países, y el *Open Data Barometer* (W3C, 2017), que se detuvo en 2015.

Recomendaciones sobre publicación de datos

El Subgrupo de Trabajo (SGT) de Política de Datos¹⁰ del Grupo de Trabajo de la Infraestructura de Datos Espaciales de España (GT IDEE)¹¹ ha realizado varios estudios a lo largo de los últimos años, cuyas conclusiones y resultados han servido como base para este artículo.

Uno de los resultados de sus trabajos es que se ha comprobado que en España hay un número notable de productores oficiales de datos geográficos, que publican esa información sin indicar ni directa ni indirectamente cuáles son las condiciones de uso ni las licencias correspondientes, lo que constituye una mala práctica y genera inseguridad jurídica en los usuarios, que desconocen cuáles son los usos permitidos y los no permitidos de los datos publicados. Esa práctica puede generar dos situaciones: que haya usuarios que realicen usos no deseados por el productor

¹⁰ <https://www.idee.es/resources/documentos/SubgruposDeTrabajo.pdf>

¹¹ <https://www.idee.es/grupo-trabajo-idee>

sin ser conscientes de ello y que otros usuarios no se atrevan a realizar algunos usos permitidos por prudencia y por querer evitar un conflicto *a posteriori* con la Administración.

Sintetizando las conclusiones y resultados, junto a la experiencia adquirida, de los trabajos de (SGT de Políticas de datos, 2019), se pueden formular algunas recomendaciones a tener en consideración a la hora de publicar datos geográficos y en general datos de cualquier tipo:

1. Adoptar de manera consciente, responsable, consistente y sostenible en el tiempo una política de datos clara, asumida por la organización, con las máximas garantías de que se mantenga y no haya pasos atrás en la apertura de datos, algo que los usuarios no suelen ni aceptar ni entender.
2. Darle la máxima publicidad posible, para que sea bien conocida, en páginas web, blog, congresos, publicaciones, redes sociales y otros canales de difusión, en los metadatos de los conjuntos de datos y en las especificaciones del producto de datos, a ser posible en español (o en el idioma autóctono) y en inglés. Lo ideal es que el usuario que se descargue los datos de manera interactiva, pase al menos una vez por una página web, en la que se le informe de la licencia de los datos.
3. Utilizar licencias de aceptación implícita, en las que el mero uso o acceso a los datos implique su aceptación, sin necesidad de tener que hacer clic en una pantalla, algo que puede ser una barrera y que no nos encontramos al leer un libro, ver una película o escuchar un tema protegidos con *copyright*.
4. Usar licencias tipo, es decir licencias estándar, de amplio uso y descripción completa en la web, sin modificaciones ni cláusulas adicionales. Proceder así posibilita la interoperabilidad de licencias, ya que si se integran dos conjuntos de datos con licencias diferentes, solo es posible saber con seguridad y de modo inmediato cómo se mezclan esas licencias si se trata de licencias tipo o estándar.
5. Si se exige el reconocimiento de la propiedad de los derechos de autor, incluir la fórmula deseada junto a la licencia o condiciones de uso.
6. Por último, siempre que sea posible, publicar datos como datos abiertos.

CONCLUSIONES

La norma UNE 148004:2018 constituye un intento de solución al problema expuesto de carencia de una descripción técnica, precisa y rigurosa, acompañada de un conjunto de pruebas abstractas que permitan la verificación y eventualmente la certificación, de qué son exactamente los datos geográficos abiertos. Es una solución restringida geográficamente, puesto que es una norma española, y temáticamente, ya que es aplicable tan solo a datos geográficos, pero es fácilmente extrapolable a otros campos de aplicación y otras latitudes, simplemente eliminado el requisito de documentar el Sistema de Referencia de Coordenadas. Se trata de una propuesta de solución consensuada y completa, que puede perfectamente modificarse de acuerdo con el consenso de una comunidad más amplia que la española y cuyo fin último no es mantener una línea de opinión determinada, sino apuntar al consenso más amplio y universal posible para tratar de solucionar un problema

que creemos sinceramente es un obstáculo para el progreso y a la rentabilización de buena parte de las ventajas que puede traer la globalización entendida en un amplio sentido.

AGRADECIMIENTOS

Agradecemos a los miembros del SGT de Trabajo de Políticas de Datos y licencias del GT IDEE, a los vocales del UNE/CTN 148 y a los compañeros, tanto del CNIG como del IGN, toda su colaboración, su apoyo y sus aportaciones, sin las que este artículo no habría sido posible.

REFERENCIAS

- Arthur, Ch. (23 de agosto de 2013). *Tech giants may be huge, but nothing matches big data*. *The Guardian*. Recuperado de <https://bit.ly/2XoXw1K>.
- Berners-Lee, T. (27 de julio de 2006). *Linked Data*. Recuperado en <https://bit.ly/3emAnEd>
- Comisión Europea (23 de marzo de 2017). *European Interoperability Framework –Implementation Strategy*. Recuperado en <https://bit.ly/3c635HI>
- Hernández, L. Nunes de Lima, V. y Smith, R. S. (19 de septiembre de 2018). *Good practices for licenses – overcoming usage barriers for INSPIRE data*. Recuperado de <https://bit.ly/2TGWI05>
- Herzog, T. (2 de noviembre de 2017). *Five years of investments in open data*. Recuperado de <https://bit.ly/3enPgWG>.
- Irving, F. (18 de julio de 2007). *We need an Open Service Definition*. Recuperado en <https://bit.ly/2Bbja2f>
- ISO (2007). ISO 19131:2017 *Geographic information – Data Product Specifications*.
- ISO (2013). ISO 19157:2013 *Geographic information – Data quality*.
- ISO (2014). ISO 19115-1 *Geographic information – Metadata. Part 1 Fundamentals*.
- Mention, Anne-Laure (ed.). (2019). *Digital Innovation: Harnessing The Value Of Open Data*. Singapur; World Scientific Publishing.
- Ministerio de Fomento. (31 de marzo de 2008). Orden Ministerial FOM/956/2008 por la que se aprueba la política de difusión pública de la información geográfica generada por el Instituto Geográfico Nacional. Recuperado de <https://bit.ly/3etzqKu>
- Ministerio de Fomento 2015. (18 de diciembre de 2015). Orden Ministerial FOM/2807/2015 por la que se aprueba la política de difusión pública de la información geográfica generada por el Instituto Geográfico Nacional. Recuperado de <https://bit.ly/3c9jkUC>
- Naciones Unidas (24 de octubre de 2018). Declaración de Dubái. Recuperado de <https://bit.ly/2yAWgAl>
- ODC. (29 de octubre de 2015). *Open Data Charter*. Recuperado de <https://bit.ly/2ZGC6QM>
- Open Society Institute (2 de diciembre de 2001). *Budapest Open Access Initiative*. Recuperado de <https://bit.ly/3d7RdGw>
- Open Knowledge Foundation. (2008). *Open Software Service Definition*. Recuperado en <https://bit.ly/2XaKlTy>

- Open Knowledge Foundation*. (2016). Definición de Conocimiento Abierto v2.1.
- Open Knowledge Foundation*. (2017). *Global Open Data Index*. Recuperado de <https://bit.ly/3equfLh>
- PIRA (30 de octubre de 2000). *Commercial Exploitation of Europe's Public Sector Information. Final Report for the European Commission*. Recuperado en <https://bit.ly/3c7nj3U>
- Real Decreto 4/2010. (29 de enero de 2010). Esquema Nacional de Interoperabilidad. Recuperado en <https://bit.ly/3dldqRu>
- Rodríguez, A.F. y Vivas, P. (19 de octubre de 2018). Informe del SGT de Políticas de Datos. Recuperado de <https://bit.ly/2ZEmOMc>
- GT Política de Datos. (25 de octubre de 2019). Panorama y evolución de las políticas de datos en la IDEE. *Reunión del GT IDEE en Cáceres*. Recuperado en <https://bit.ly/36CLki7>
- Singleton, D. A. et al. (5 de febrero de 2016). *Establishing a Framework for Open Geographic Information science*. Recuperado en <https://bit.ly/3bZqQkJ>
- The Economist* (6 de mayo de 2017). *The world's most valuable resource is no longer oil, but data*. Recuperado de <https://econ.st/3d3JnxT>
- UK Cabinet Office. (18 de junio de 2013). *Open Data Charter*. Recuperado de <https://bit.ly/2X5PbkK>
- UNE (2005). UNE 66175:2003 Sistemas de gestión de la calidad. Guía para la implantación de sistemas de indicadores.
- UNE (2018). UNE 148004:2018 Datos geográficos abiertos.
- UNECE (1998). Convención sobre el acceso a la información, la participación del público en la toma de decisiones y el acceso a la justicia en asuntos ambientales. Recuperado en <https://bit.ly/2TFwKRE>
- Unión Europea (31 de diciembre de 2003). Directiva 2003/98/CE relativa a la reutilización de la información del sector público. Recuperado en <https://bit.ly/3goCI3k>
- Unión Europea (26 de junio de 2013). Directiva 2013/37/UE relativa a la reutilización de la información del sector público. Recuperado en <https://bit.ly/36vop8w>
- Unión Europea (4 de mayo de 2016). Reglamento (UE) 2016/679 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de esos datos. Recuperado de <https://bit.ly/3c4AbYG>
- Unión Europea (20 de junio de 2019). Directiva (UE) 2019/1024 relativa a los datos abiertos y la reutilización de la información del sector público. Recuperado en <https://bit.ly/2XxbiQ3>
- Valentino-De Vries et al. (10 de diciembre de 2018). *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*. Recuperado en <https://nyti.ms/3gouN5U>
- W3C (*World Wide Web Consortium*). (Mayo de 2017). *Open Data Barometer. Global Report 4th edition*. Recuperado de <https://bit.ly/2ZFedZO>

