
ESTUDIOS / RESEARCH STUDIES

El uso de las licencias libres en los datos públicos abiertos

Luis Fernando Ramos-Simón*

*Facultad de Documentación. Universidad Complutense de Madrid
Correo-e: lframoss@ucm.es | ORCID iD: <http://orcid.org/0000-0003-2267-8405>

Recibido: 29-02-2016; 2ª versión: 19-09-2016; 3ª versión: 04-11-2016; Aceptado: 08-11-2016.

Cómo citar este artículo/Citation: Ramos-Simón, L. F. (2017). El uso de las licencias libres en los datos públicos abiertos. *Revista Española de Documentación Científica*, 40(3): e179. doi: <http://dx.doi.org/10.3989/redc.2017.3.1376>

Resumen: Los datos masivos o big data recopilados a través de distintas herramientas tecnológicas o extraídos de grandes colecciones de información en distintos formatos son capaces de generar nuevo conocimiento en los sectores más diversos, pero, a la vez, plantean problemas tanto de titularidad como de su uso posterior. El estudio analiza el concepto de análisis de datos, los distintos tipos de datos y su posible regulación, presentando un panorama sobre cómo en los últimos años los países más avanzados en apertura de datos han comenzado a aplicar unos esquemas de licencias abiertas que remueven los obstáculos legales para incorporar estos tipos de datos a la creación de nuevo conocimiento. La investigación realizada muestra cómo distintos gobiernos y otras corporaciones y organismos internacionales productores de datos públicos han encontrado en estas licencias —en particular, las modalidades PPDL y CC0 en las que se renuncia a todos los derechos sobre las bases de datos— un mecanismo adecuado para favorecer el uso y reutilización de estos grandes volúmenes de datos.

Palabras clave: Acceso abierto; datos públicos abiertos, licencias de información; derechos de autor; análisis de datos; derecho sui generis sobre bases de datos; minería de datos; datos masivos; reutilización de información del sector público (ISP); Creative Commons; Opendatacommons.

The use of free licenses in open public data

Abstract: *Big data*, the massive amounts of data collected using different technological tools or extracted from large collections of information in different formats, are able to generate new knowledge in diverse sectors. But problems are posed regarding both their ownership and subsequent use. This study analyzes the concept of data analysis, as well as the different types of data and their possible regulation, presenting an overview of how in recent years some open licensing schemes have begun to be implemented for avoiding the legal obstacles facing the inclusion of these data types in the creation of new knowledge. The study shows how different governments and international organizations that produce public data have found an appropriate mechanism for promoting the use and reuse of large volumes of public data through these licenses - in particular, the PPDL and CC0 modalities in which all rights to databases are waived.

Keywords: Open access; open governmental data; information licenses; copyright; data analysis; sui generis database rights; data mining; big data; reuse of public sector information (PSI); Creative Commons; Opendatacommons.

Copyright: © 2017 CSIC. Este es un artículo de acceso abierto distribuido bajo los términos de la licencia *Creative Commons Attribution (CC BY)* España 3.0.

1. INTRODUCCIÓN

El uso de los datos se intensifica cada año. Estos datos sirven para ayudar en soluciones tan diversas como en el diagnóstico de daños cerebrales, a encontrar el lugar ideal para la instalación de parques eólicos, a evitar la congestión del tráfico o a prever el rendimiento de los cultivos en países en desarrollo (Comisión Europea, 2014). Al tiempo aumentan los dispositivos de recogida de datos en ámbitos públicos y privados (calles, comercios, estaciones, aeropuertos, carreteras...) y el uso de redes de datos (meteorológicos, espaciales, cultivos, fauna...), así como las redes sociales y controles de navegación en Internet proporcionan un enorme caudal de datos, fácilmente convertibles en conocimiento en distintos contextos (Fitzgerald y otros, 2009). Además, actividades como los proyectos de digitalización de colecciones bibliotecarias han puesto de manifiesto el gran valor que tienen esas páginas de texto escaneadas, digitalizadas y categorizadas después de haber sido convertidas en datos, valor que va mucho más allá de poner al alcance de la humanidad grandes colecciones bibliotecarias en dominio público. Todo ello ha convertido a los datos en la gran revolución de nuestro tiempo (Mayer-Schönberger y Cukier, 2013).

La eclosión de estos datos masivos recopilados a través de diferentes métodos y herramientas tecnológicas (tratamiento masivo de textos y datos de todo tipo, sensores, satélites, cámaras...) da lugar a nuevas actividades mediante las cuales se analiza la información resultante para obtener nuevo conocimiento de todo tipo (McKinsey Global Institute, 2013), lo que plantea problemas de legalidad referidos a la titularidad de los propios datos y cuestionan si socialmente tales datos deben seguir reglas de propiedad o más bien deben disponer de un estatuto abierto que permita un uso generalizado de los mismos. El estudio muestra cómo, frente a otras alternativas, las licencias se han convertido en un instrumento relativamente sencillo y práctico plenamente adecuado para satisfacer las expectativas de usuarios y reutilizadores. De este modo, en el ámbito de los portales gubernamentales de los países calificados con una mejor apertura de datos, las licencias libres estándar internacionales, ya sea adoptándolas directamente o bien haciéndolas compatibles con las licencias gubernamentales gozan de una casi plena aceptación como licencias por defecto, siendo las que más favorecen el libre uso de las bases de datos, tales como PDDL y CC0, las que parecen adaptarse mejor a ese entorno de datos públicos abiertos. A pesar de todo, todavía queda mucho camino por recorrer en este propósito de poner los datos fácilmente accesibles para todos, libres e interoperables.

El contenido del artículo está estructurado en las siguientes partes:

- El concepto de análisis de contenido o minería de datos.
- Protección legal de los datos.
- Aplicación de las licencias libres a los datos públicos abiertos.
- Conclusiones.

2. OBJETIVOS Y METODOLOGÍA

Los objetivos fundamentales de la investigación son los siguientes:

- Analizar los nuevos conceptos y técnicas de tratamiento de los datos mediante los cuales se obtienen nuevas posibilidades de uso y reutilización para generar nueva información y conocimiento.
- Estudiar las características básicas de las bases de datos y sus formas de protección por derechos de autor. La revolución de los datos pone de relieve la necesidad de abordar transformaciones en su marco regulatorio (en particular, en las bases de datos y las licencias), de modo que las bases de datos sean fácilmente consultables para cualquier propósito.
- Determinar las licencias que mejor se adecúan a la puesta a disposición de los datos abiertos organizados para su reutilización en forma de bases de datos.
- Examinar los tipos de licencias que emplean los portales gubernamentales mejor considerados a nivel internacional y las condiciones de reutilización de los datos públicos accesibles.
- Contrastar y comparar el uso de las licencias estándar y compatibles para determinar que licencias se adaptan mejor a la oferta de datos públicos abiertos.

Con este propósito, se ha llevado a cabo una pormenorizada búsqueda en los portales gubernamentales de once países (los diez primeros más España, 13^o) que están entre los más avanzados en la apertura de datos, conforme a los resultados del Open Data Barometer (WWW Foundation, 2015), en este indicador se analizan y comparan parámetros de unos noventa países. Sus resultados se presentan en dos cuadros comparativos en el que se analizan las licencias empleadas —por defecto y compatibles— en la puesta a disposición de los datos públicos abiertos. El estudio revela que los portales gubernamentales de los países más avanza-

dos en la gestión de los datos abiertos disponen de licencias que permiten la reutilización de todo tipo de bases de datos mediante una regulación propia que además es compatible con las licencias estándar internacionales más avanzadas orientadas al libre flujo de los datos sin restricciones.

Además del apartado empírico, este trabajo también se nutre de la experiencia obtenida en diversos proyectos, así como de estudios y trabajos sobre activos de información en el ámbito de los datos y contenidos abiertos llevados a cabo por el grupo de investigación Publidoc-UCM desde 2006 (MHAP, 2012). Desde esta perspectiva temporal, llama la atención que se ha pasado de una situación inicial de tímida apertura de datos con situaciones muy diversas y declaraciones de principios muy generales, al escenario actual en el que una licencia de datos abiertos por defecto es el eje central que convive con otros usos diferenciados, ya sea regulando el uso mediante pago con tarifas o bien distinguiendo uso comercial de no comercial.

3. DESARROLLO

3.1. El concepto de análisis de datos o minería de contenidos

La minería de datos es el término más tradicional para designar el uso de tecnología y arquitectura diseñadas para detectar información procesable y así obtener rendimiento económico y social del gran volumen de información depositado en bases de datos y que es accesible directamente o través de Internet, ámbito conocido como datos masivos o Big Data¹, para lo cual se utilizan herramientas para la recopilación, extracción y recombinación de distintos elementos con objeto de obtener nuevos conocimientos.

Frente a un concepto más clásico que se centraba sólo en la minería sobre los datos, las nuevas capacidades tecnológicas permiten incluir textos, imágenes, figuras y todo tipo de materiales digitales o elementos documentales, lo cual ha conducido a una pérdida de valor del concepto original y al surgimiento de denominaciones alternativas a minería de datos, tales como minería de textos, minería de contenidos, análisis de datos...

De tal modo que, en primer lugar debemos sistematizar los elementos que integran el concepto que ahora se conoce como análisis de datos o minería de contenidos:

La minería de datos es una técnica computacional que permite cruzar variables en un entorno informático para descubrir nuevo conocimiento a partir de la información contenida en las bases de datos. De una forma más precisa, "la minería de datos

es un conjunto de técnicas, métodos y algoritmos que permiten extraer información útil desde fuentes de distinto tipo", señala Mendoza (2011). Los métodos de la minería de datos (DM, Data Mining por sus siglas en inglés) "permiten descubrir esta información y transformarla en un valioso conocimiento tanto retrospectivo (histórico) como prospectivo (proyecciones) o comprensivo (entender lo que ocurre)" (Colle, 2002). Se emplean distintos métodos y modelos para la minería de datos, siendo el KDD (Knowledge Discovery in Databases) el primer modelo aceptado por la comunidad científica en 1996. A partir del 2000 surgen tres nuevos modelos el SEMMA (iniciales en inglés de Muestreo, exploración, modificación, modelado y valoración), Catalyst (P3TQ, producto, mercado, precio, tiempo y cantidad) y CRISP-DM, siendo este modelo el más aceptado actualmente, estructurando la minería de datos en fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación e implantación (Moine y otros, 2011).

En cuanto a la tipología de los datos, Hand y otros (2001) distinguen entre datos observacionales y datos experimentales para separar los datos que son recopilados sin más propósito que registrar algún hecho, mientras que los datos experimentales son resultado de la actividad programada, de ahí que desde su perspectiva, "la minería de datos es el análisis de (a menudo grandes) conjuntos de datos observacionales para encontrar relaciones desconocidas y resumir los datos en nuevas formas que sean a la vez comprensibles y útiles para el titular de los datos". Más recientemente, Melero y Hernández-San Miguel (2014), en el ámbito de los datos de investigación añaden otras dos categorías: datos computacionales (datos de entrada y derivados de ciertos programas) y datos de simulación (generados a partir de modelos de prueba).

Por su parte, la minería de textos es una actividad de más reciente aparición, surge con los procesos de digitalización masiva iniciados en 2004 con el proyecto Google Books, razón por lo que el término ha recibido menos atención que la minería de datos (Truyens y Van Eecke, 2014). Además de texto entran en el objeto de análisis las imágenes, vídeos, sonidos, figuras... La peculiaridad de esta nueva forma de análisis de los textos radica en que los procesos de indexación, tratamiento y/o uso de los textos e imágenes pueden estar más próximos a las aspiraciones de los autores por lograr una protección o retribución por este tratamiento que es subsiguiente al proceso creativo, como veremos más adelante.

Algunos autores (Triaille y otros, 2014) para estudiar este nuevo fenómeno, prefieren la expresión

análisis de datos, en lugar de minería de textos y datos, que definen como "el procesamiento automático de materiales digitales, los cuales pueden incluir textos, datos, sonidos, imágenes u otros elementos o una combinación de estos, con el fin de descubrir nuevo conocimiento o ideas estratégicas". El análisis de datos es un concepto más amplio que minería que se asocia a la extracción de información, mientras que el análisis implica "rastreo, procesamiento, comparación, copia, análisis, recuperación, interpretación, búsqueda, sintaxis, etc."

Por su parte, la expresión minería de contenidos es la preferida por otros autores (Guadamuz y Cabell, 2013), para definir el proceso de análisis de información mediante software que busca patrones interesantes o importantes que no podrían ser observados de otros modos, en ámbitos tales como los hospitales, el medio ambiente, el comercio o los sentimientos sociales. El análisis de contenido ha crecido debido al incremento de la capacidad de computación centrado en el descubrimiento de nuevo conocimiento en las grandes bases de datos, ámbito en el que la minería de datos sería una de sus ramas.

Como se observa, esta nueva actividad es el resultado de la capacidad de análisis que proporcionan los grandes sistemas de información, haciendo posible la obtención de nuevo conocimiento a partir del análisis de grandes volúmenes de datos. Desde el punto de vista legal, este nuevo fenómeno debe de ser estudiado en el marco regulatorio de las bases de datos. Es decir, las recopilaciones de obras, de datos o de otros elementos independientes dispuestos de manera sistemática o metódica y accesibles individualmente por medios electrónicos o de otra forma, es la realidad sobre la que se asienta toda esta actividad de análisis de datos que es nuestro objeto de estudio.

3.2. Tipos de datos y su protección como base de datos

En sí mismo, el dato es un componente que aislado no tiene ningún significado (Mendo y otros, 2013), pero "organizado en conjuntos de datos o bases de datos pueden constituir un activo de gran valor en actividades como la investigación, el marketing, la economía, y la gestión en todos los ámbitos: comercial, transporte, recursos humanos, hospitales..."

Como resulta evidente, todos los datos para su tratamiento y gestión automatizada están organizados en conjuntos o bases de datos, aspecto esencial para determinar las posibilidades de acceso y uso de la información allí contenida. La legislación europea (Parlamento Europeo, 1996), señala que "tendrán la consideración de «base de datos» las

recopilaciones de obras, de datos o de otros elementos independientes dispuestos de manera sistemática o metódica y accesibles individualmente por medios electrónicos o de otra forma" (art. 1.2).

Asimismo, están protegidas las bases de datos que no cumplen el requisito de la originalidad, pero en las que se ha realizado una inversión sustancial, ya sea financiera, de esfuerzo o de tiempo, para la "obtención, la verificación o la presentación de dicho contenido"(art. 7), es lo que la Directiva europea denomina bases de datos *sui generis* y que ha creado una notable controversia (Correa, 2007) por los efectos que tiene sobre la protección y apropiación del conocimiento en general, dado que si gran parte del conocimiento y memoria mundial está almacenada en bases de datos, todos esos contenidos pertenecerían a los titulares de esas bases de datos, con independencia de que sus contenidos sean creaciones originales, datos de dominio público o simples datos o representaciones obtenidas de la naturaleza.

De acuerdo a la definición y al contenido de la información de dichas bases de datos, podemos establecer las cuatro categorías que se especifican a continuación, si bien una misma base de datos puede tener protección jurídica en una o varias de esas categorías, aspecto a tener en cuenta en cuanto a su acceso, consulta y uso. Es especialmente relevante que las condiciones de la licencia seleccionada se adecúen al uso o reutilización pretendidos.

3.2.1. Bases de datos originales

Como decíamos más arriba, la protección de los derechos de autor se refiere a su estructura en cuanto forma de expresión o disposición de los contenidos, sin alcanzar a estos. Lo que se protege es la selección o disposición de tales contenidos, con independencia de éstos, ya sea por su diseño, formato o la sistemática de las anotaciones u ordenación de los elementos. Asimismo, los contenidos pueden estar también protegidos, en el caso de que por sí mismos constituyen obras protegibles (por ejemplo, un artículo académico incorporado a una base de datos científica). Sólo podrá consultarse sin autorización del autor con fines de ilustración de la enseñanza o de investigación científica, en la medida justificada, sin fines comerciales e indicando la fuente.

3.2.2. Bases de datos amparadas por un derecho "sui generis"

El segundo tipo de bases de datos mencionado en la Directiva de 1996, se refiere a las bases de datos "no originales" (López, 2002) en las que su

titular ha realizado una inversión sustancial para la "obtención, verificación o presentación de su contenido" (art. 7). Es importante subrayar que si bien ambas bases de datos (originales y amparadas por el derecho *sui generis*) están protegidas legalmente, la inclusión en uno u otro apartado es determinante, por cuanto en un caso se protege la obra como un todo, en el segundo caso, sólo la inversión sustancial y durante un período de quince años, si no se actualiza la base de datos (Cámara-Lapuente, 2007).

Como se ha señalado desde el lado latinoamericano, al igual que desde otros países y posiciones, el principal argumento en contra de la protección de las bases de datos fácticas no es que proteja las bases de datos en sí mismas, sino "la información en ellas contenida, con el consiguiente riesgo de que se establezcan límites a la circulación de aquella, incluso de la que hasta el presente se mantiene en el dominio público" (López, 2002). Añade el autor, en este estudio bien documentado, que hay otros dos argumentos más en contra de la protección, de un lado genera protecciones perpetuas —haciendo imposible el paso al dominio público— al ser su caducidad renovable y favorece la creación de mecanismos de rastreo y recopilación, en lugar de incentivar el esfuerzo creativo. Otros análisis de la directiva europea de bases de datos sugieren que la idea de dotar de una ventaja competitiva a los productores de bases de datos descuidó el necesario equilibrio de intereses entre editores, usuarios y autores (Bitton, 2008).

Además de estas críticas en contra del monopolio que suponen estas bases de datos, ha surgido también un movimiento igualmente poderoso a favor del libre uso de la información bajo el presupuesto de que su uso abierto y generalizado fomenta el bienestar social y crea riqueza.

3.2.3. Bases de datos públicos

Los datos públicos existentes en organismos públicos de la Unión Europea tienen un tratamiento específico en virtud de las Directivas de reutilización de información del sector público de 2003 y 2013, en las que se prevé la puesta a disposición del público para la reutilización de sus documentos "cualquier contenido sea cual sea el soporte (escrito en papel o almacenado en forma electrónica o como grabación sonora, visual o audiovisual —art. 2.3—)" conservados por organismos del sector público para fines comerciales o no comerciales. El principio general es la libre puesta a disposición de tales contenidos, en caso de aplicar alguna tarifa, ésta "se limitará a los costes marginales en que se incurra para su reproducción, puesta a disposi-

ción y difusión" (art. 6.1)². La mayoría de los países europeos están incorporando ambas directivas mediante la creación de portales de datos abiertos y los correspondientes mecanismos de licencias a través de los cuales se da acceso a las bases de datos públicas.

La información del sector público es tradicionalmente el principal proveedor de información sistematizada, estructurada y de buena calidad en la mayoría de los sectores (información geoespacial, meteorológica, catastral, etc.). Desde la década de los noventa del siglo XX esta información fue objeto de atención especial ante la visión de que en una sociedad basada en la información, la generada en el sector pública iba a ser una materia prima esencial para la innovación y el crecimiento económico (Comisión Europea, 1998).

Entonces, el panorama era muy distinto a un lado y otro del Atlántico, mientras que en Estados Unidos la Ley de Libertad de Información (U.S. Department of Justice, 2016) proclama desde 1966 el libre acceso de todos los ciudadanos a la información generada por el Gobierno Federal no existiendo derechos reservados sobre la información administrativa federal (Comisión Europea, 1998)³. Sin embargo, Europa hubo de hacer frente a una realidad muy fragmentada en cuanto a tradiciones administrativas, lenguas y territorios. Esa era la situación cuando se aprueba en 2003 la Directiva de información del sector público con el propósito de crear unas mínimas reglas comunes en cuanto a transparencia, igualdad de acceso a los documentos y al principio de tarifas basadas en los costes. La Directiva de 2013 (Parlamento Europeo, 2013), actualización de la de 2003, ha profundizado principalmente en la apertura de datos libres o a coste marginal, en el fomento de la reutilización y en la estandarización de formatos. Estas iniciativas y la política de datos abiertos del comienzo de la presidencia de Obama han impulsado en todo el mundo los portales de datos nacionales desde donde son accesibles miles de recursos en formatos de base de datos, aunque todavía queda un gran camino en cuanto a la actualización de datos, formatos e interoperabilidad, aspectos centrales para llevar a cabo actividades eficientes de análisis de datos.

El procedimiento que se pretende generalizar es el de acceso a través de licencias abiertas para permitir el libre uso de los datos, es decir, "el dato es legalmente abierto cuando o bien está en dominio público o con una licencia que permite el uso y reutilización comercial y no comercial sin restricciones", tal como señala la Guía de datos abiertos de la ONU (2013). A pesar de los avances que se dan en la implantación de licencias abiertas y a lo que luego nos referiremos, es necesario subrayar que la proli-

feración de condiciones de uso de los datos, la falta total de información de cómo usarlos y el empleo de licencias que no son adecuadas aparecen como las principales barreras de acceso a la información del sector público, tal como se ha señalado reiteradamente en los últimos años, así en el informe de la consultora Deloitte, encargado por el Gobierno del Reino Unido (Deloitte, 2013) o en el estudio de Vickery (2012) para la Comisión Europea.

3.2.4. Bases de datos personales

Estas bases de datos, como su nombre indica, contienen datos referidos a las personas⁴ y, por lo tanto, están sometidas a la legislación específica de protección de datos personales (Parlamento Europeo, 1995)⁵. La protección de los derechos de autor no puede "hacerse extensiva a su contenido y se entenderá sin perjuicio de los derechos que pudieran subsistir sobre dicho contenido" (Parlamento Europeo: 1996). En este sentido, el uso de bases de datos personales para fines distintos a los que fueron creadas exige un proceso previo de anonimización de los datos personales para cumplir la Directiva en cuanto a la libre circulación de esos datos personales. Es decir, estas bases de datos cuando los datos sean anonimizados tendrán la protección que les corresponda en función de sus características. Un tratamiento específico, pero que no reviste ninguna particularidad respecto a los derechos de autor, reciben las bases de datos de solvencia patrimonial y de prospección comercial (Rebollo y Serrano, 2014).

4. ACCESO Y USO DE LOS DATOS: EXCEPCIONES A LOS DERECHOS DE AUTOR Y LICENCIAS SOBRE LOS DATOS

Como queda explicado, el acceso, uso y posible reutilización de la información contenida en las bases de datos está regulado por las leyes, así como mediante contratos y/o licencias. Las licencias de información⁶ son una modalidad que se ha generalizado en Internet por ser un elemento sencillo para modular el acceso a la información y en particular por ofrecer mecanismos simples para regular las relaciones entre autores y usuarios de contenidos protegidos por derechos de autor, de hecho es la solución impulsada por la legislación europea ("Los Estados miembros alentarán a todos los organismos del sector público a que utilicen las licencias modelo", art. 8.2, Parlamento Europeo, 2013). Un aspecto interesante es si una licencia o contrato puede prevalecer sobre lo dispuesto en un texto legal, disponiendo así, por ejemplo, si puede o no ser usado un material para realizar minería de datos o textos, aspecto que no tiene una solución unánime en las legislaciones europeas (Truyens y

Van Eecke, 2014). Mayor interés para nuestro propósito reviste el debate sobre si es posible aplicar una excepción a los derechos de autor al análisis de los datos. Veamos a continuación este aspecto, antes de adentrarnos en las licencias para uso de datos públicos.

4.1. Las excepciones de derechos de autor en el análisis de datos

El debate sobre el acceso y uso de la información en Internet, es decir, qué tipo de protección de derechos de autor deben tener los contenidos, se ha extendido también a si deben estar protegidas las bases de datos en las actividades de minería de datos. Las posiciones no son muy diferentes a las que existen en torno al acceso abierto y al acceso comercial en los contenidos protegibles por derechos de autor; sin embargo, en el aspecto concreto del análisis de datos parece abrirse paso una excepción a los derechos de autor en favor de permitir ciertos usos, al menos los que tienen por finalidad la investigación.

Por lo descrito hasta ahora resulta difícil encontrar una base de datos cuyo acceso no pueda ser restringido por su titular, pues aun en los casos en que sean contenidos que están en dominio público o datos del sector público, sus recopiladores pueden ejercer la titularidad en base a la inversión económica efectuada, baste pensar en las colecciones digitales como Google Libros, u otras muchas colecciones digitales de las bibliotecas en donde aparece con frecuencia el símbolo de protección de los derechos de autor (©). De modo más rotundo, Tuyens y Van Eecke (2014) aseveran que fuera de la investigación no comercial es difícil crear un estatuto en favor de estas actividades de análisis de datos, especialmente en los casos de bases de datos comerciales y con textos archivados que tengan un mínimo nivel de creatividad. En este sentido, los autores comentan la diferencia entre Europa y Estados Unidos en donde no existe una legislación comparable sobre bases de datos y se aplican de modo más flexible las reglas del "fair use", esto es, un criterio interpretativo que se adapta a los cambios socioeconómicos y tecnológicos sin tener que actualizar constantemente la legislación formal.

Una cuestión clave en el debate sobre el análisis de datos es si los autores deberían estar autorizados a prohibir el uso de una obra para un propósito diferente al de la creación original, por ejemplo, en los casos en que 100 horas de grabaciones o de mil novelas que sean tratadas para analizar aspectos tales como la imagen de la mujer, el grado de uso de una tecnología o el vocabulario o la sintaxis empleada en un grupo social. La solución

es difícil que sea concluyente, en un primer momento la respuesta debería de ser negativa porque los derechos de autor protegen la expresión creativa, mientras que el análisis de datos de cada obra constituye una materia prima (*raw data*) que sólo adquiere valor junto con muchos otros textos. Por contra, la posición a favor de la protección reforzará el papel de quienes se aprovechan de la creatividad ajena a través de la mera acumulación de datos. Un aspecto no menor es si en tales actos de explotación la autorización le correspondería a cada autor por su aporte creativo o al titular de la base de datos, puesto que el bien explotado sería la mera colección de los datos. El asunto lejos de ser quimérico podría afectar a miles de editores, autores, colaboradores periodísticos, intérpretes, guionistas, compositores...

En este contexto se ha abierto un debate en Europa y en otras partes del mundo sobre si es posible la introducción de una excepción legal en el ámbito de la propiedad intelectual en las actividades de análisis de datos asegurando un equilibrio de los diferentes intereses implicados. La excepción ya está en aplicación o en estudio en algunos países (Japón o Reino Unido).

Un estudio muy sólido encargado para la Comisión Europea (Jean Paul Triaille y otros, 2014) aborda un análisis pormenorizado sobre el uso del análisis de datos en las bases de datos. En síntesis, tanto las bases de datos originales como las protegidas con derecho "sui generis", desde un punto de vista legal están protegidas, por lo tanto es necesario abrir vías alternativas para permitir el uso de las técnicas de análisis de datos, y que incluso la aplicación de la excepción de investigación —reconocida en la Directiva de bases de datos— encuentra dificultades interpretativas y, además, en dicha excepción se hace explícito que no está permitido el uso comercial. Así las cosas, los autores del informe proponen varias alternativas: facilitar acuerdos de entendimiento entre titulares y usuarios de las bases de datos, adoptar un documento interpretativo en donde la Comisión Europea clarifique los aspectos de la Directiva que dificultan una interpretación armonizada o proponer un cambio legislativo en el que se introduzca una excepción en toda la Unión Europea para facilitar las técnicas de análisis de datos, siendo ésta la vía por la que se inclina el documento⁷ y la que ha respaldado la Comisión Europea (2016).

5. APLICACIÓN DE LAS LICENCIAS LIBRES EN LOS DATOS PÚBLICOS ABIERTOS

La introducción de una excepción legal para permitir una utilización menos restringida de los datos ofre-

cería seguridad, sin embargo, esa solución es lenta, más en el entorno de la Unión Europea, en donde a la aprobación de las directivas hay que sumar otro período de aplicación en cada país. Por el contrario, las licencias de información ofrecen una solución sencilla en la circulación por Internet de contenidos protegidos por derechos de autor. La aparición en 2003 de Creative Commons y en 2004 de Open Knowledge Foundation (Murray-Rust, 2008), han servido para impulsar las licencias abiertas, llamadas también libres, porque son capaces de indicar a personas no expertas, mediante unos pocos símbolos, los derechos, permisos y restricciones que ofrece el titular de los derechos a sus posibles usuarios, lo que facilita la circulación por Internet de contenidos sin tener que adaptarse previamente a la singularidad de cada país y de su sistema jurídico. En el caso de los datos se han abierto vías para que algunas modalidades de licencias permitan cualquier tipo de uso o reutilización, al incluir estas licencias la renuncia a todos los derechos por parte de los titulares, incluido el derecho sobre las bases de datos no originales.

5.1. Las licencias libres aplicables a los datos

Como es sabido, mediante estas licencias abiertas o libres (Liang, 2004), (Rosen, 2005), los titulares de contenidos protegidos por derechos de autor limitan algunas formas de explotación de la obra (explotación comercial, creación de obras derivadas, reutilización...) y permiten otros usos (usos comerciales o no, compartir con otros las modificaciones, etc.). Estas condiciones de explotación se incorporan a la obra mediante un código digital al difundirla en Internet, de modo que el posible lector o usuario de la obra en línea conoce de antemano cuáles son los derechos que el autor le permite o no, sin demasiada ambigüedad. Desde hace años estas licencias se usan en algunas áreas de conocimiento (O'Boyle y otros, 2011), en gran parte de publicaciones científicas en acceso abierto, así como en otros sitios en los que se acogen contenidos producidos por los usuarios, tan populares como *Youtube* o *Flickr*, y cada vez más lo hacen algunos gobiernos con la información pública puesta a disposición de los reutilizadores, siendo esta práctica considerada razonablemente apropiada en el ámbito académico y comienzan a sugerir que las licencias vayan incorporadas a los "datasets" para facilitar su uso (Allen, 2014).

Señaladamente, en el ámbito del tratamiento de los datos científicos, un gran número de organizaciones del sector han optado por poner los datos y bases de datos a libre disposición de la comunidad científica mediante este tipo de licencias (Hrynaszkiewicz y Cockerill, 2012), adecuadas al

tratamiento de los datos, que pretenden soslayar las diferencias entre las legislaciones nacionales, con el objetivo central de que los posibles derechos de autor no actúen como una barrera en el progreso del conocimiento. En sectores donde predominan los datos públicos, como el de los geodatos se aboga claramente en favor de que los titulares de derechos promuevan un uso de los datos sin restricciones creando una comunidad de información (Information Commons) que beneficie a usuarios y titulares (CODATA, 2015), (GITA White Paper, 2005). No obstante, un estudio señala que estamos todavía lejos de ese propósito al mostrar que de 4.012 bases de datos analizadas, la mitad no informaba sobre las condiciones de uso del sitio (Guadamuz y Cabell, 2013). Según esta fuente, de las bases de datos con licencias de acceso, las utilizadas con más frecuencia forman parte de las familias de licencias Creative Commons (no se detalla por separado la licencia específica para datos u metadatos CC0, si bien en la fecha del estudio esta licencia estaba poco implantada) y de Open Data Commons. En Francia, parece también que ambas licencias son las más utilizadas, aunque el autor percibe que los datos se usan poco y se hace preciso crear unas condiciones propicias para impulsar el uso de los datos públicos (Boustany, 2013).

A continuación, reseñamos las características más relevantes de estos dos grupos de licencias en el uso de los datos abiertos:

Dentro del conjunto de licencias de Creative Commons (Ramos, 2014)⁸, la licencia CC0 1.0 Universal (CC0 1.0) (http://creativecommons.org/publicdomain/zero/1.0/deed.es_ES), está especialmente orientada al uso de datos y metadatos. De hecho, los más de cincuenta millones de metadatos de los objetos digitales de Europea son accesibles bajo esa licencia. La licencia CC0, como se indica en su título, es un "ofrecimiento al dominio público". De este modo el proveedor de los datos renuncia a todos los derechos de propiedad intelectual (incluido el derecho *sui generis* sobre las bases de datos) que recaigan sobre los datos suministrados; ello implica que el proveedor ha de obtener autorización de terceros en el caso de que no sea titular de los mismos. Asimismo, si la renuncia de derechos del proveedor sobre los metadatos no es legalmente posible a través de la licencia CC0, será de aplicación la licencia pública alternativa (Public License Fallback), lo que implica que otorga a cada persona afectada "una licencia exenta de cánones, intransferible, no sublicenciable, no exclusiva, irrevocable e incondicional para el ejercicio" de todos los derechos sobre la obra en todo el mundo y por la duración máxima prevista en la ley aplicable. En la actualidad, la página Web de Creative Commons

(https://wiki.creativecommons.org/CC0_use_for_data) relaciona 49 organizaciones que usan la licencia CC0, tan relevantes como la Organización Europea para la Investigación Nuclear (CERN), Europea o U.S. Government Data, el portal de datos del Gobierno de Estados Unidos.

En el ámbito de las licencias Creative Commons los otros dos tipos de licencia que se usan con frecuencia en la apertura de datos son la licencia CC-BY (Reconocimiento) y CC-SA (Compartir igual). En la primera, el titular de los datos exige que se reconozca su titularidad. En la segunda, CC-SA, el titular permite reutilizar y compartir los datos con tal de que la distribución de esos datos resultantes se haga con la misma licencia. Estas dos licencias no son específicas para compartir bases de datos *sui generis*, aunque siguen siendo las más utilizadas para estos propósitos.

La familia de licencias Open Data Commons (ODC), fue creada en 2006 a partir de un borrador elaborado por Talis Platform (Miller; Styles; Heath, 2008) están promovidas por la organización Open Knowledge Foundation (OKF) (<http://opendatacommons.org/licenses/>) que dispone también de tres tipos licencias de datos abiertos para su aplicación a las bases de datos:

- Public Domain Dedication and License (PDDL) — "Public Domain for data/databases". La licencia es, como la licencia CC0, una cesión al dominio público, mediante la cual el titular de los derechos de autor ofrece la obra al beneficio público y renuncia a los derechos de autor y derechos *sui generis* que puedan recaer sobre la base de datos, por lo cual la obra es libre y está abierta para otros usos. Asimismo, la renuncia es para todo el mundo y para todos los derechos presentes y futuros en cualquier formato. Del mismo modo, la licencia prevé que si la renuncia resulta inaplicable en alguna jurisdicción, el licenciador autoriza todos los usos sobre la base de datos mediante una licencia no exclusiva, libre de derechos en todo el mundo y por todo el tiempo de protección de la obra. Además, la licencia contiene una renuncia expresa al ejercicio de los derechos morales.
- Attribution License (ODC-By) — "Attribution for data/databases". La licencia permite a los usuarios compartir libremente, modificar y usar la base de datos con el único requisito de atribuir la paternidad de la obra, libre de derechos, no exclusiva por todo el tiempo que dure la protección de la base de datos, con derecho de extracción y reutilización de la base de datos. La base de datos puede

estar protegida por otros derechos que no están contenidos en la licencia (datos personales, contratos privados o marcas...), asimismo recoge la renuncia a los derechos morales del titular. Es similar a la CC-BY (Reconocimiento de la autoría).

- Open Database License (ODC-ODbL) — “Attribution Share-Alike for data/databases”. La licencia es la tercera de las opciones y su finalidad es permitir a los usuarios compartir libremente, modificar y usar la base de datos manteniendo esa misma libertad para otros usuarios, es decir, con cláusula “copyleft”. Por lo demás, la licencia es similar a la anterior en cuanto cobertura temporal y derechos de acceso y uso de la base de datos, no afectando los contenidos individualmente. Esta licencia también es muy parecida a la licencia CC-SA, de modo que las obras derivadas deben disponer de la misma licencia para compartir las obras.

5.2. Resultados: uso de las licencias en los portales de datos públicos

Al igual que en el ámbito científico, en el sector público se toma conciencia de que la información es una fuente de progreso y riqueza, de ahí la necesidad de poner a disposición de los ciudadanos y de las empresas la información y datos generados por el sector público. Si la Unión Europea, como mencionamos arriba, ya había adoptado en 2003 la Directiva de reutilización de información del sector público (modificada en 2013), la OCDE adoptó sendos documentos para impulsar el uso de datos abiertos en todos los países miembros (OCDE, 2007 y 2008), al proponer una mejora del acceso y uso de la información del sector público para que se adoptan unos principios en tal sentido.

Interesa subrayar la recomendación que incide en favorecer las condiciones de acceso y transparencia en la reutilización, “haciendo disponible y desarrollando un sistema de licencias automatizadas en línea para dar cobertura a la reutilización cuando se aplican licencias, teniendo en cuenta los derechos de autor”. En este sentido se recomienda favorecer la reutilización para facilitar su acceso y uso, incluyendo mecanismos de licencias sencillas para hacer las obras ampliamente accesibles al público.

De hecho, a partir de 2009, coincidiendo con la declaración de Obama sobre datos abiertos (The Obama Administration’s... 2011), se han dado grandes avances en la apertura de datos públicos. Muchos gobiernos y organizaciones internacionales han puesto grandes colecciones de datos a disposición del público que, o bien se acogen a una licencia

estándar que permite un amplio uso de los datos, o adaptan su propias licencias nacionales⁹ a los estándares de licencias promovidos por organizaciones como la OKF (Ramos y otros, 2012). En todas estas iniciativas de datos abiertos se asumen unos principios básicos sobre los que se debe publicar la información y configurar las licencias. Estos principios (ONU, 2013, 37) son:

- Información accesible en Internet.
- En formatos procesables electrónicamente.
- Licencia abierta que permita la reutilización comercial.
- Libre de cargas y sin restricciones de reutilización.

De tal modo que las Licencias para un “Gobierno abierto” se asientan sobre estos principios básicos “Open Government”, tal como la impulsada por el gobierno británico (HM Government, 2012) en las que se establece:

- Libertad para copiar, publicar, distribuir y transmitir información.
- Adaptar la información.
- Hacer uso comercial, combinarla con otras informaciones o incluirla en los propios productos o aplicaciones.
- Con la obligación básica de incluir la fuente de información y la licencia empleada.

En este contexto, en la Tabla I, se muestran las licencias de referencia, por defecto, que utilizan los portales gubernamentales de datos de once países: los diez países mejor clasificados a nivel mundial (y España, 13ª posición) en el ranking Open Data Barometer (WWW Foundation, 2015).

A pesar de la variedad de situaciones se aprecia una clara tendencia en el predominio de la aplicación de las licencias libres cuyas características reseñamos en el epígrafe anterior. Aunque todavía falte mucho camino por recorrer en la homogeneización de los requisitos para el acceso a los datos y su reutilización —como demuestra con mucha precisión Morando (2013) en un estudio sobre interoperabilidad legal— debe quedar claro que la aplicación de estas licencias supone un gran avance en el panorama internacional de intercambio y uso de los datos.

Algunos informes nacionales de los organismos de datos abiertos, como el de Holanda, reconocen el liderazgo indiscutible que ejercen en esta materia Reino Unido y Estados Unidos, por cuanto hacen disponibles los datos en un portal nacional lo cual resulta adecuado para un gobierno abierto y transparente. En esta cuestión sobrevuela el de-

Tabla I. Licencias de referencia en los portales gubernamentales de los países con mejor apertura de datos

Puesto	País	Web Portal de datos	Licencia
1	Reino Unido	http://data.gov.uk/	<i>Open Government Licence</i> , por defecto: Recomendada para todos los organismos públicos que no aplican cargos. http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/ . Compatible con CC-BY 4.0 y ODC-BY.
2	Estados Unidos	http://www.data.gov/about	Los datos federales se ofrecen libres y sin restricciones. Los datos no federales pueden tener diferentes licencias. http://www.data.gov/privacy-policy#license . Recomendada atribución: <i>Data retrieved from Data.gov</i> (https://www.data.gov/)
3	Suecia	http://data.riksdagen.se/In-English/	Los datos públicos sobre el Gobierno están disponibles en la web Riksdag, "sus bases de datos están libremente disponibles para cualquiera que desea usarlas" http://data.riksdagen.se/In-English/
4	Francia	https://www.data.gouv.fr/fr/	https://www.data.gouv.fr/fr/terms/ , por defecto, compatible con estándares internacionales (ODC-BY, CC-BY 2.0)
5	Nueva Zelanda	http://www.data.govt.nz/	<i>Creative Commons Attribution 3.0</i> (CC-BY3.0), por defecto para los contenidos que están en el sitio https://www.data.govt.nz/copyright/ . Los datasets enlazados desde el portal se rigen por la licencia que establezca cada agencia.
6	Holanda	http://data.overheid.nl/	(CC-BY 4.0), a menos que se indique lo contrario. La mayor parte de los contenidos son geodatos (Open Data Trend Report 2015, p. 32).
7	Canadá	http://open.canada.ca/en	<i>Open Government License-Canada</i> , muy parecida a la británica. Licencia mundial, no exclusiva para cualquier propósito. No indica expresamente la compatibilidad con CC-BY y ODC-BY.
8	Noruega	http://data.norge.no/nlod/en/1.0	Licencias compatibles: para toda la información <i>Open Government Data</i> (NLOD). Para información que no constituye bases de datos: <i>Creative Commons Attribution Licence</i> (CC-BY 3.0). Para aquella parte de información que está en base de datos: <i>Open Data Commons Attribution License (version 1.0)</i> (ODC-BY)
9	Dinamarca	http://digitaliser.dk/	Licencia estándar de datos abiertos ("standard license which corresponds to CC-0 and CC-BY"). (https://digitaliser.dk/resource/2432531/artefact/Conditions_for_use_of_Danish_public_sector_data.docx). Corresponde con las licencias CC-0 and CC-BY, dependiendo de si el organismo público exige reconocimiento o no. Aclaración y enlace facilitados por correo electrónico.
10	Australia	http://data.gov.au/	http://data.gov.au/about . <i>Creative Commons Attribution 3.0 Australia licence</i> , por defecto.
13	España	http://datos.gob.es/	http://datos.gob.es/?q=aviso-legal . Asimilable a <i>Creative Commons Attribution</i> ("siempre que se ajusten a lo establecido en este real decreto", RD 1495/2011).

Fuente: <http://www.opendatabarometer.org> (2015) y elaboración propia. Datos consultados/actualizados en octubre de 2016.

bate de si la apertura de datos y la libertad de información son dos aspectos complementarios o en competencia. Sobre este aspecto, diversos análisis realizados en el ámbito europeo ponen de relieve que la apertura de datos repercute en las solicitudes de información en base a las leyes de transparencia, ya que si los gobiernos gestionan bien la información los costes de cada petición bajarán. En el mismo sentido, cuanto mayor es el coste de un servicio público mayor resulta el impacto que produce la transparencia y apertura de datos en ese servicio, según se pone de relieve en el informe holandés (Open Data Trend Report, 2015).

Los resultados revelan algunos aspectos peculiares en cuanto a las licencias y el uso de los datos:

- La licencia del gobierno británico sobre datos abiertos es la piedra angular que ha servido de referencia a todos los gobiernos, como se reconoce ampliamente y se recoge en el cuadro.
- El número de reutilizadores de información pública es muy escaso todavía, tal como ponen en evidencia algunos datos, por ejemplo, en Francia.

- Asimismo, la oferta de datos públicos, aunque crece a buen ritmo, es desigual y la calidad de los mismos es mediana (Francia, Holanda, Dinamarca, España). Como se pone de manifiesto en la excelente página de ayuda que contiene el portal francés, a veces se deben extraer los datos de documentos cerrados (en pdf o de tabulaciones en páginas web).
- Hay muchos organismos públicos que no ofrecen datos abiertos, como también se desprende de los datos manejados.
- En el aspecto positivo, con respecto a hace unos años (Ramos y otros, 2012) hay un notable incremento de buenas prácticas en aspectos tales como la oferta de tutoriales para transformar los datos, grandes mejoras en el uso de herramientas de visualización, así como para la conversión de datos de unos formatos a otros.
- Cada vez son más populares los encuentros para avanzar en la mejora de los datos y su reutilización (*hackatones, opendata camps*) de carácter festivo y competitivo que suelen durar un fin de semana (Francia, Reino Unido, España...).

Como se observa en la Tabla II, la mayoría dispone de una licencia propia elaborada por cada Gobierno que son muy parecidas entre sí (Reino Unido, Francia, Noruega, Dinamarca, Australia, Nueva Zelanda, incluso España), que es compatible con el libre uso de los datos o reconoce la compatibilidad de las condiciones de uso de la información pública con alguna de las licencias libres. Sirva como modelo el portal francés (data.gouv.fr) donde se contextualiza la licencia abierta en el marco de las licencias estándar internacionales y cita expresamente la del gobierno británico y las licencias estándar compatibles (ODC-BY, CC-BY 2.0). Asimismo, esta interpretación es la que cabe hacer en el marco de aplicación de estas licencias a la reutilización de la información del sector público en España. Como acertadamente señalan Clabo y Ramos (2015), estas licencias se adaptan "a cada uno de los usos definidos en la normativa española de reutilización y pueden servir para precisar y acotar el alcance de las condiciones fijadas en el aviso legal o para precisar los distintos usos de aquellos datos y documentos".

Parece interesante señalar la diferencia entre las versiones 3.0 y 4.0 de la licencia Creative Commons. Frente a las versiones anteriores que permitían adaptaciones nacionales y generaban confu-

Tabla II. Modalidades y compatibilidad de licencias libres empleadas en los portales gubernamentales seleccionados

Licencia por defecto	Países	Compatibilidad entre licencias	Volumen de datos en el portal
Sin licencia	Estados Unidos Suecia	-CC-0 y PDDL -(https://theunitedstates.io/licensing/) Datos libres de pago y de licencias	192.261 datasets 300.000 documentos aprox.
CC0 y CC-BY	Dinamarca	-Atribución del titular: CC-BY -Sin Atribución: CC-0	n.d.
CC-BY 4.0	Holanda	CC-BY 4.0	9.052 datasets
OGL	Reino Unido Francia Canadá España	CC-BY 4.0 y ODC-BY ODC-BY y CC-BY 2.0 No explícita No explícita	40.067 datasets 19.558 datasets 121.844 registros 12.021 resultados
CC-BY 3.0	Nueva Zelanda Noruega Australia	No se especifica -No base de datos: CC-BY 3.0 -Sí base de datos: ODC-BY No se especifica	4.387 datasets 757 datasets 23.371 datasets

Fuente: Elaboración propia a partir de los portales de datos mencionados en el cuadro I (octubre de 2016).

sión sobre si quedaba licenciado el uso de las bases de datos no originales, la versión 4.0, lanzada en 2013, considera expresamente incluidos en la licencia tanto la información como los datos protegidos por derechos de autor o por el derecho sui generis sobre bases de datos vigente en la Unión Europea y otros países (<https://creativecommons.org/licenses/by/4.0/legalcode>). Esa es la causa por la que la licencia Noruega excluye la compatibilidad de la licencia Creative Commons para el uso de los datos. Con la versión CC 4.0 se ha tratado de evitar el efecto de las llamadas licencias "portables", adaptadas a cada jurisdicción nacional con diferentes matices lo que impedía una interpretación uniforme en aspectos tan importantes como el de si se consideraban incluidos dentro de la licencia los derechos sobre las bases de datos no originales.

Mención aparte merece la puesta a disposición de la información y datos que hacen los portales de los gobiernos estadounidense y sueco. En la página dedicada al uso de los datos se pone muy de relieve que los datos del Gobierno de Estados Unidos se ofrecen, en dominio público, sin ninguna reserva de derechos, situación preferible a la información licenciada. Al abordar la compatibilidad se indica que las licencias CC0 y PDDL son muy similares, ambas son una renuncia a los derechos sobre los contenidos y allí donde no sea de aplicación la renuncia, permite una plena reutilización. En cambio, "*open licensing* presupone una protección por derechos de autor". Por su parte, en el portal sueco sólo se recoge la mención de que los datos abiertos pueden ser utilizados libre de cargos y sin necesitar licencias.

6. CONCLUSIONES

La investigación pone de relieve que los portales gubernamentales adoptan licencias por defecto compatibles con declaraciones internacionales sobre gobierno abierto o licencias libres estándar de uso internacional (Creative Commons y Open Data Commons) que permiten acceder y usar los datos en distintas modalidades con un mecanismo relativamente sencillo. Muchas de estas iniciativas asumen unos principios básicos que permiten la transformación de los datos, sin más requisitos que la cita de la fuente, e incluso se abren paso las licencias orientadas al libre flujo de datos y metadatos, más próximas al concepto de dominio público, como PDDL y CC0. El uso de estas licencias facilita el libre flujo de datos y su reutilización. Sin embargo, será necesario avanzar también en el uso de metadatos y en la expresión de derechos en Internet, puesto que en este aspecto la asignación de derechos y la combinación de contenidos (por ejemplo, que haya distinta li-

encia para los datos y el texto de un artículo), se convierte en un asunto de cierta complejidad tecnológica, pues hay que aplicar un lenguaje de derechos para cada tipo (REL, por sus iniciales en inglés). (Abelson y otros, 2012)

El concepto tradicional de minería de datos realizado sobre las bases de datos se ha extendido al análisis de textos, datos e imágenes, tratamiento que ahora se realiza sobre datos masivos o macrodatos (big data), de modo que estas actividades son conocidas como análisis de datos o análisis de contenido, fundamentales en todos los sectores sociales y económicos por cuanto son muy eficaces para analizar pautas históricas y, a partir de ahí, diseñar modelos y respuestas.

La gestión de estos grandes volúmenes de datos debe ser puesta en relación con su posible colisión con algunos derechos. Se suele afirmar que salvo en los datos en dominio público, el análisis de datos puede afectar a los derechos de los autores o de los titulares de las bases de datos no originales. Si bien los argumentos en torno a la protección de los datos por su originalidad son bastante débiles, la protección de tales bases de datos por razón de la inversión o el esfuerzo realizado supone que todas las bases de datos cuenten con algún tipo de protección legal, así sucede en todo el ámbito de la Unión Europea.

La cuestión clave en el debate sobre los datos y la autoría es dirimir si los autores deberían estar autorizados a prohibir el uso de una obra para un propósito diferente al de la creación original, en donde no se analiza la creatividad o la forma de expresión de las ideas sino, simplemente, a partir de grandes volúmenes de estos contenidos (grabaciones, novelas...) se obtienen datos de utilidad social y comercial que no están directamente relacionados con la actividad creativa. Por ahora el acceso a estos grandes volúmenes de datos muy demandados, sólo puede hacerse mediante la aprobación de una excepción a los derechos de los autores, o bien mediante el uso de licencias de acceso a dichos datos. El debate está muy condicionado por la evolución constante de la tecnología.

Por su parte, el uso de licencias de información resulta, en principio, una solución más fácil y práctica. Sin embargo, algunos estudios ponen de relieve que en muchos casos no se explicita la vía de acceso a las bases de datos y cuando sí existe no siempre la licencia resulta adecuada a los datos, por lo cual se aboga por una simplificación de las licencias y cesiones de los datos, así como por la formación de usuarios y proveedores, así como por modificar hábitos seculares (Boustany, 2013). Esa misma complejidad la hemos podido comprobar al

analizar los 11 portales gubernamentales, ya que son necesarias búsquedas pormenorizadas en cada sitio web para determinar cuál es la licencia aplicable por defecto, e incluso preguntar directamente en la institución responsable del portal de datos. Asimismo, a veces es muy tortuoso distinguir entre la licencia para usar-reutilizar las bases de datos que opera el portal y sus metadatos de la base de datos o *dataset* que contiene la referencia a los datos que se pretenden descargar o consultar.

Por último, aunque sea desde una perspectiva muy subjetiva, en el caso de los portales gubernamentales de datos abiertos se pone de relieve una clara evolución en estos últimos años desde unos términos de uso con declaraciones muy heterogéneas entre sí a la situación actual, algo más sencilla centrada en el uso de una licencia de datos abiertos por defecto. Pero no perdamos de vista que los conjuntos de datos públicos que se ofrecen a libre reutilización son relativamente pocos en algunos países, como queda claro en la tabla II. El tiempo nos dirá cuántos organismos públicos no se

acogen a la licencia por defecto de datos abiertos y a qué sectores pertenecen. El desconocimiento del impacto de los *big data* podría influir en la oferta de datos públicos.

7. AGRADECIMIENTOS

Este trabajo forma parte de los resultados del proyecto de investigación titulado "Organización del acceso, uso y reutilización de la información del sector público en España. Hacia la consolidación de una industria de la información", financiado por el Plan Nacional de I+D en España (Ref.: CSO2010-17451).

ACKNOWLEDGEMENTS

This study is part of the results of the research project entitled "Organization of access, use and reuse of public sector information in Spain. Towards the consolidation of information industry", funded by the National R + D Plan in Spain (Ref.: CSO2010-17451).

8. NOTAS

- 1 La consultora IDC (2012) define las tecnologías Big Data como una nueva generación de tecnologías y arquitecturas diseñadas para extraer valor económico en muy grandes cantidades de una amplia variedad de datos haciéndolo posible mediante captura a alta velocidad, exploración y/o análisis. Por su parte, Ubaldi (2013), en el estudio de la OCDE sobre datos públicos abiertos, define "Big data" como toda colección de *datasets* tan grande y compleja que es difícil usar herramientas tradicionales de gestión de bases de datos para su procesamiento lo que incluye captura, búsqueda, compartición, transferencia, análisis y visualización.
- 2 La Directiva 2013 ha sido incorporada al Derecho interno español mediante la Ley 18/2015, de 9 de julio, por la que se modifica la Ley 37/2007, de 16 de noviembre, sobre reutilización de información del sector público (BOE, 10/07/2015). En el texto legal se prevé la aplicación de tarifas por el suministro de documentos para su reutilización, "limitándose la misma a los costes marginales en que se incurra para su reproducción, puesta a disposición o difusión", literalmente igual que en el texto de la Directiva.
- 3 La Ley de Libertad de Información de Estados Unidos (FOIA, por sus siglas en inglés), ha tenido varias modificaciones desde su aprobación inicial en 1966. El texto actualizado con las modificaciones de 2016 está disponible en <https://www.justice.gov/oip/freedom-information-act-5-usc-552> (última consulta, 12/10/2016).
- 4 La Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, en su art. 2 a) entiende por a) «datos personales»: toda información sobre una persona física identificada o identificable (el «interesado»); se considerará identificable toda persona cuya identidad pueda determinarse, directa o indirectamente, en particular mediante un número de identificación o uno o varios elementos específicos, característicos de su identidad física, fisiológica, psíquica, económica, cultural o social;
- 5 La mencionada Directiva 95/46/CE, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos, establece que en los Estados miembros no se podrá ni restringir ni prohibir el uso de los datos personales por la razón de garantizar el derecho a la intimidad en lo que respecta al tratamiento de los datos personales.
- 6 Preferimos este término "licencias de información" en lugar de recursos de información, para referirnos al amplio ámbito de licencias que regulan contenidos informativos sujetos a derechos de autor intercambiados a través de Internet, así lo utiliza en español el profesor Fernández Molina, véanse las referencias bibliográficas correspondientes.
- 7 Según se ha anunciado en septiembre de 2016, esta propuesta de facilitar el acceso a los datos de investi-

gación parece abrirse camino en la Unión Europea (ver propuesta de Directiva (COM (2016)593), presentada en septiembre de 2016: <http://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52016PC0593&rid=1>). La Comisión Europea propone entre las reformas a los derechos de autor, la creación de una excepción, sólo para organizaciones científicas "actuando en interés público", sólo para los contenidos a los que se haya accedido legalmente y cubriría la investigación comercial y no comercial, pero no se aplicaría a las firmas comerciales (*Nature News*, 15/09/2016).

- 8 Remitimos a los apartados del capítulo de esta obra sobre Europea en el que se explican las licencias para acceder al contenido y los metadatos accesibles desde Europea. Al igual que sucede en muchos portales de datos abiertos, es necesario distinguir los metadatos que contiene la base de datos del portal respecto de los contenidos a los que enlaza y que pertenecen a distin-

tas instituciones, con lo cual las condiciones de acceso pueden ser diferentes de unas bases de datos a otras e incluso entre las bases de datos accesibles dentro de la misma institución. Todo lo cual hace compleja la descarga y uso de los datos.

- 9 Ese ha sido el camino seguido por varios países. El primero fue el Reino Unido, país pionero en el uso de licencias en línea por la OPSI (Office Public Sector Information) hasta su desaparición y que después adoptó la licencia Open Government Licensing para ofrecer sus datos públicos, la cual es esencialmente equivalente a CC-BY u ODC Attribution. Esta misma práctica fue seguida por Francia e Italia que también adoptaron los principios de esas licencias libres a su propia licencia nacional (Morando, 2013, 443). España siguió ese mismo camino con la aprobación del RD1495/2011, en vigor, que regula la reutilización de los datos públicos del Gobierno central.

9. REFERENCIAS

- Abelson, H.; Adida, B.; Linksvayer, M.; Yergler, Y. (2012). CC REL: The Creative Commons Rights Expression Language. En: Dulong de Rosnay, M.; Martin, J.C. de (eds.). *The Digital Public Domain*. OpenBook Publishers; Cambridge (R.U.). <http://www.openbookpublishers.com/product/93/the-digital-public-domain-foundations-for-an-open-culture> [último acceso, 04/09/2016].
- Allen, B. (2014). *Access to and use of public sector information: the academic re-user perspective*. Gobierno de Australia: Information policy reports. <https://www.oaic.gov.au/resources/information-policy/information-policy-resources/access-to-and-use-of-public-sector-information-the-academic-re-user-perspective.pdf> [último acceso, 04/09/2016].
- Bitton, M. (2008). Exploring European Union Copyright Policy through the Lens of the Database Directive. *Berkeley Technology Law Journal*, 23 (4), 1411-1470. <http://dx.doi.org/doi:10.15779/Z38TH7W>
- Boustany, J. (2013). Accès et réutilisation des données publiques État des lieux en France. *Les Cahiers du numérique*, 9 (1), 21-37. <https://doi.org/10.3166/lcn.9.1.9-19>
- Cámara-Lapuente, S. (2007). Derecho sui generis sobre las bases de datos. En: Rodríguez Tapia, J.M. (Dir.). *Comentarios a la Ley de Propiedad Intelectual*. Editorial Thomson-Civitas; Cizur Menor (Na).
- Clabo, N.; Ramos, I. (2015). Reutilización de datos abiertos en la administración pública en España y uso de licencias-tipo. *Revista Española de Documentación Científica*, 38(3): e097. <http://dx.doi.org/10.3989/redc.2015.3.1206>
- CODATA (2015) *The Value of Open Data Sharing*. https://www.earthobservations.org/documents/dsp/20151130_the_value_of_open_data_sharing.pdf [último acceso, 20/10/2016].
- Colle, R. (2002). *Explotar la información noticiosa. Data mining aplicado a la documentación periodística*. Dpto. Biblioteconomía y Documentación de la Universidad Complutense, Madrid.
- Comisión Europea (2016). Propuesta de Directiva del Parlamento Europeo y del Consejo sobre los derechos de autor en el mercado único digital. Documento COM 2016 (593).
- Comisión Europea (Comunicado 2/7/2014). *La Comisión insta a los gobiernos a aprovechar el potencial de los macrodatos*. http://europa.eu/rapid/press-release_IP-14-769_es.htm [último acceso, 04/09/2016].
- 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*. Documento COM 1998 (585). http://cordis.europa.eu/pub/econtent/docs/gp_es.pdf [último acceso 12/09/2016].
- Correa, C. (2007) *Protección del dominio público: acceso a datos e información científica y su utilización*. Informe de Comité Nacional de Ética en la Ciencia y la Tecnología (CECTE). Argentina.
- Deloitte (2013). *Market Assessment of Public Sector Information*. Department for Business, Innovation and Skills; Londres. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/198905/bis-13-743-market-assessment-of-public-sector-information.pdf [último acceso, 19/10/2016].
- Dulong de Rosnay, M.; Janssen, K. (2014). Legal and Institutional Challenges for Opening Data across Public Sectors: Towards Common Policy Solutions. *Journal*

- of *Theoretical and Applied Electronic Commerce Research*, 9 (3), 1-14. <https://doi.org/10.4067/s0718-18762014000300002>
- Fernández-Molina, J.C. (2004a). Licensing Agreements for Information Resources and Copyright Limitations and Exceptions. *Journal of Information Science*, 30 (4), 337-346. <https://doi.org/10.1177/0165551504045853>
- Fernández-Molina, J.C. (2004b). Los límites al derecho de autor en las licencias de información electrónica: la necesidad de una reforma legal. *Actas do Congresso Nacional de Bibliotecários, Arquivistas e Documentalistas*, nº 8.
- Fitzgerald, A.M.; Fitzgerald, B.F; Pappalardo, K.M. (2009). The Future of Data Policy. En: Hey, T.; Tansley, S; Tolle, K. (Eds.). *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, pp. 201-208. <http://eprints.qut.edu.au/28182/> [último acceso, 04/09/2016].
- GITA White Paper (2005). *Free or Fee: The Governmental Data Ownership Debate*. <http://www.calmapping.org/conferences/2005/ccma2005-GITA-free-or-fee.pdf> [último acceso, 04/09/2016]
- Guadamuz, A.; Cabell, D. (2014). Data mining in UK higher education institutions: Law and policy. *Queen Mary Journal of Intellectual Property*, 4 (1): 3-29. <https://doi.org/10.4337/qmjip.2014.01.01>
- Hand, D.; Mannila, H; Smyth, P. (2001). *Principles of Data Mining*. Massachusetts, USA, The MIT Press Cambridge.
- HM Government (2012). *Open Data White Paper | Unleashing the potential*. http://data.gov.uk/sites/default/files/Open_data_White_Paper.pdf [último acceso, 04/09/2016].
- Hrynaszkiewicz, I.; Cockerill, M. (2012) Open by default: a proposed copyright license and waiver agreement for open access research and data in peer-reviewed journal. *BMC Research Notes* 2012, 5:1. <https://doi.org/10.1186/1756-0500-5-494>
- IDC/Comisión Europea (2012). *Worldwide Big Data Technology and Services*. <https://ec.europa.eu/digital-single-market/en/news/worldwide-big-data-technology-and-services-2012-2015-forecast> [último acceso, 04/09/2016].
- Liang, L. (2004). *The Guide to Open Content Licenses V.1.2*. Piet Zwart Institute; Amsterdam.
- López, A. (2002). *El Impacto de la Protección de las Bases de Datos no Originales sobre los Países de América Latina y el Caribe*. Comité permanente de derechos de autor y derechos conexos. Organización Mundial de la Propiedad Intelectual (OMPI).
- Mayer-Schönberger, V.; Cukier, K. (2013). *Big data. La revolución de los datos masivos*. Turner Publicaciones, Madrid.
- McKinsey Global Institute (2013) *Open data: Unlocking innovation and performance with liquid information*. McKinsey & Company. <http://www.mckinsey.com/business-functions/business-technology/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information> [último acceso, 04/09/2016].
- Melero, R.; Hernández-San-Miguel, J. (2014). Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. *Revista Española de Documentación Científica*, 37(4): e066. <https://doi.org/10.3989/redc.2014.4.1154>
- Mendo Carmona, C.; Ramos Simón, L. F.; Arquero Avilés, R.; Valle Gastaminza, F. del; Botezán, I.; Sánchez, R.; Tejada, C.; Peón, J.L.; Cobo, S.; Sala, A. (2013). Del acceso a la reutilización, del dato al documento: una visión conceptual de la información pública. *Revista Española de Documentación Científica* 36(3): e013. <https://doi.org/10.3989/redc.2013.3.957>
- Mendoza Rocha, M. (2011). Minería de datos en la Web. En: Cacheda Seijo, F.; Fernández Luna, J.M.; Huete Guadix, M., *Recuperación de información. Un enfoque práctico y multidisciplinar*. Ra-Ma; Madrid.
- MHAP (Ministerio de Hacienda y Administraciones Públicas). (2012). *Funcionamiento, Mantenimiento y Actualización del Catálogo de Información Pública. Datos. gob.es*. MHAP. Centro de Publicaciones; Madrid.
- Miller, P.; Styles, R.; Heath, T. (2008). Open Data Commons. A License for Open Data. <http://ceur-ws.org/Vol-369/paper08.pdf> [último acceso, 04/09/2016].
- Moine, J.M.; Haedo, A.S.; Gordillo, S.E. (2011). Estudio comparativo de metodologías para Minería de Datos. En: *XIII Workshop de Investigadores en Ciencias de la Computación*. <http://hdl.handle.net/10915/20034>
- Morando, F. (2013). Legal Interoperability: Making Open (Government) Data Compatible with Businesses and Communities. *JLIS.it*. Vol.4, n.1 (Gennaio/January 2013).
- Murray-Rust, P. (2008). Open data in Science. *Serials Review*, 34 (1), 52-64. <https://doi.org/10.1080/00987913.2008.10765152>
- O'Boyle, N. M.; Guha, R.; Willighagen, E. L.; Adams, S. E.; Alvarsson, J.; Bradley, J.C.; Murray-Rust, P. (2011). Open Data, Open Source and Open Standards in chemistry: The Blue Obelisk five years on". *Journal of Cheminformatics*, 3(37). <https://doi.org/10.1186/1758-2946-3-37>
- OECD/OCDE (2007). *OECD Principles and Guidelines for Access to Research Data from Public Funding*. <http://www.oecd.org/science/sci-tech/38500813.pdf> [último acceso, 04/09/2016].
- OECD/OCDE (2008). *OECD Recommendation of the council for enhanced access and more effective use of public sector information*. [C (2008)36]. <http://www.oecd.org/sti/44384673.pdf> [último acceso, 04/09/2016].
- Open Knowledge Foundation (2012). *Open Data Handbook*. <http://opendatahandbook.org/guide/es/> [último acceso 04/09/2016].
- Open Data Trend Report (2015). Informe presentado al Parlamento de Holanda por *Algemene Rekenkamer* (Tri-

- bunal de Cuentas). http://www.courtfoaudit.nl/english/Publications/Audits/Introductions/2015/03/Open_Data_Trend_Report_2015 [último acceso, 14/10/2016].
- Organización de las Naciones Unidas (ONU), (2013). *Guidelines on open government data for citizen engagement*. <http://workspace.unpan.org/sites/Internet/Documents/Guidelines%20on%20OGDCE%20May17%202013.pdf> [último acceso, 04/09/2016].
- Parlamento Europeo (1995). Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. *Documento DOUE-L-1995-81678*. <http://www.boe.es/buscar/doc.php?id=DOUE-L-1995-81678> [último acceso, 04/09/2016].
- Parlamento Europeo (1996). Directiva 96/9/CE del Parlamento Europeo y del Consejo, de 11 de marzo de 1996, sobre la protección jurídica de las bases de datos. *DOUE-L-1996-80413*. <http://www.boe.es/buscar/doc.php?id=DOUE-L-1996-80413> [último acceso, 04/09/2016].
- 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. *DOUE-L-2013-81251*. https://www.boe.es/diario_boe/txt.php?id=DOUE-L-2013-81251 [último acceso, 04/09/2016].
- Ramos Simón, L.F. (2014). Declaraciones de Derechos sobre los metadatos y contenidos de Europeana. En Ramos Simón, L.F.; Arquero Avilés, R. (coords.). *Europeana. La plataforma del patrimonio cultural europeo*. Trea. Gijón.
- Ramos Simón, L.F.; Arquero Avilés, R.; Botezán, I.; Cobo Serrano, S.; Sala Jiménez, A.; Sánchez Jiménez, R.; Valle Gastaminza, F. del (2012). De la reutilización de la información del sector público a los portales de datos abiertos en Europa. *BiD: textos universitaris de biblioteconomia i documentació*, núm. 29. <http://bid.ub.edu/29/ramos2.htm> [último acceso 15/10/2016].
- Rebollo, L.; Serrano, M. (2014). *Manual de Protección de datos*. Dykinson, Madrid.
- Rosen, L. (2005). *Open Source Licensing: Software Freedom and Intellectual Property Law*. Upper Saddle River, New Jersey: Prentice Hall Professional Technical Reference.
- The Obama Administration's commitment to Open Government: a status report (2011). https://obamawhitehouse.archives.gov/sites/default/files/opengov_report.pdf
- Triaille, J.P.; Meeûs d'Argenteuil, J.; Francquen, A. (2014). *Study on the legal framework of text and data mining (TDM)*. Unión Europea. http://ec.europa.eu/internal_market/copyright/docs/studies/1403_study2_en.pdf [último acceso 04/09/2016].
- Truyens, M.; Van Eecke, P. (2014). Legal aspects of text mining. *Computer law & security review* 30 (2) 153-170. <https://doi.org/10.1016/j.clsr.2014.01.009>
- Ubaldi, B. (2013). Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives. *OECD Working Papers on Public Governance*, No. 22, OECD. <https://doi.org/10.1787/5k46bj4f03s7-en>
- U.S Department of Justice (2016). The Freedom of Information Act 5 U.S.C.§552. <https://www.justice.gov/oip/freedom-information-act-5-usc-552> [último acceso, 8/11/2016].
- Vickery, G. (2012). Review of recent studies on PSI reuse and related market developments. París: Information Economics. Informe para la Comisión Europea (ejemplar mecanografiado). <https://ec.europa.eu/digital-agenda/en/news/review-recent-studies-psi-reuse-and-related-market-developments> [último acceso, 04/09/2016].
- WWW (World Wide Web Foundation) (2015). *Open Data Barometer Global Report (Second Edition), 2015*. <http://opendatabarometer.org/2ndEdition/> [último acceso 04/09/2016].