

D.4. El que tiene los datos tiene el poder

Por Jesús Tramullas

3 enero 2011

Tramullas, Jesús. "El que tiene los datos, tiene el poder".
Anuario ThinkEPI, 2011, v. 5, pp. 117-120.



Resumen: El aumento de servicios de información en internet que hacen un uso intensivo de datos de todo tipo, como por ejemplo las redes sociales, plantea problemas sobre la custodia y propiedad de los mismos, así como problemas legales de protección de datos y privacidad. La tensión sobre la propiedad y la utilización de los datos se está extendiendo a todos los ámbitos, y no es ajena tampoco al mundo de bibliotecas y archivos.

Palabras clave: Silos de datos, Redes sociales, Propiedad de los datos.

Title: *They who have the data have the power*

Abstract: The development of online information services that make intensive use of data and information of all kinds, such as social networks, raises questions about their custody and ownership. It also raises legal issues of data protection and privacy. The tension over ownership and use of data is spreading to all areas, and is no stranger either to the world of libraries and archives. This text reflects on this problem.

Keywords: Data silos, Social networks, Data ownership.

VIVIMOS UN MOMENTO TECNOLÓGICO en el cual los usuarios tienden a identificar la interfaz que utilizan con el sistema o aplicación que está por debajo de la misma, y en el que las aplicaciones se simplifican hasta extremos insospechados hace pocos años.

En este contexto existe una peligrosa tendencia a olvidar un principio básico y fundamental, aceptado y aplicado en el diseño de aplicaciones informáticas desde hace más de treinta años, y es que de una cuestión técnica se está transformando en una cuestión social, política y económica.

1. El principio de independencia de los datos

Este principio establece que los datos deben ser independientes de las aplicaciones que los manejan. Viene del campo del diseño de bases de datos relacionales, y ya fue incluido por **Codd** en sus doce reglas.

http://es.wikipedia.org/wiki/12_reglas_de_Codd

De una manera más cercana y coloquial, es la independencia entre las aplicaciones y los ficheros que contienen los datos (la información) la que hace posible acciones tan sencillas como hacer copias de seguridad, intercambiar ficheros o exportar e importar entre diferentes formatos, y permite acciones más complejas como reutilizar

información entre aplicaciones y servicios web de diferentes proveedores.

El acceso libre a la información almacenada demanda la existencia de estándares que establezcan las características lógicas y físicas de los ficheros que la almacenan, como es el caso de *OpenDocument* o *Acrobat pdf*.

"El interés de servicios como Facebook, Twitter, Slideshare o Flickr no radica en las funciones que ofrecen sino en la base de datos a la que dan acceso"

2. Internet y los silos de datos

Un silo de datos es un gran depósito de información estructurada o semiestructurada. Hasta el advenimiento de internet la mayoría eran privados y de difícil acceso. Con la popularización de la Red comenzó a estar disponible una mayor cantidad de información de los silos para los usuarios, mucha de ella gratuitamente, otra mediante pago. La proliferación de las tecnologías y aplicaciones de la web social a comienzos de la década de 2000 hizo que comenzaran



<http://www.freebase.com/view/government>

a crearse nuevos silos, cuyo contenido se ha ido formando con información proveniente de dos fuentes principales:

- comportamientos y acciones que los usuarios llevan a cabo en la Red;
- contenidos e información que los usuarios han aportado a la propia Red.

Puede imaginarse la gran cantidad de información de ambos tipos que atesoran servicios como *Facebook*, *Twitter*, *Slideshare* o *Flickr*, por citar unos pocos. El interés de servicios como los citados no radica en las funciones que ofrecen, sino en la base de datos a la que dan acceso. *Facebook* vale lo que vale no por sus funciones de comunicación, que realmente son básicas, comunes y conocidas de antes, sino por la red de datos de usuarios a la que se puede tener acceso. En realidad, por los datos y la información de otros usuarios a los que ofrece acceso con la metáfora de red social.

3. El que tiene los datos tiene el poder

En un contexto como el actual, en el que el desarrollo de aplicaciones se ha abaratado y acelerado, la riqueza no está en las propias aplicaciones. Quien tiene el poder es quien tiene el silo de datos, que es el que puede imponer políticas de usos permisibles y no permisibles de los datos contenidos en el mismo.

Muchas de las aplicaciones innovadoras que alcanzan la cima del efímero éxito "2.0" se fundamentan en que aprovechan o explotan los silos de datos que ofrecen otros. *Google*, siempre tan avizor, se ha posicionado antes que la compe-

tencia en el campo de la explotación de los silos de datos comprando *Metaweb* y su producto estrella *Freebase*.

<http://googleblog.blogspot.com/2010/07/deeper-understanding-with-metaweb.html>

En el campo de la gestión de información, un modelo parecido ya se vivió durante las décadas de los 80 y 90, cuando los *hosts* de bases de datos campaban a sus anchas con sus tarifas y servicios. Esos tiempos pasaron, evidentemente, pero no hay que descartar que a medio plazo muchos silos de datos gratuitos vayan pasando

a modelos mixtos para rentabilizar sus costes de mantenimiento y explotación.

“El uso de los silos de datos se ha vuelto motivo de fricción entre proveedores y usuarios del mundo de la información”

No son muchos los usuarios que se toman la molestia, necesaria molestia, de leer los términos de uso de los servicios que contratan. Deberían hacerlo. Es necesario aclarar que cuando se habla de "propiedad de los datos" los usuarios tienden a pensar que los datos son suyos y pueden hacer con ellos lo que quieran, y es cierto... hasta cierto punto.

Por ejemplo, las condiciones de uso de *Facebook* establecen en su declaración de derechos y responsabilidades, punto 2: "eres el propietario de todo el contenido y la información que publicas en *Facebook*, y puedes controlar cómo se comparte..." hasta que se llega al punto 2.1: "...nos concedes una licencia no exclusiva, transferible, con posibilidad de ser sub-otorgada, sin *royalties*, aplicable globalmente, para utilizar cualquier contenido de PI que publiques en *Facebook* o en conexión con *Facebook*" (PI: contenido protegido por derechos de propiedad intelectual, como fotografías y vídeos).

<http://www.facebook.com/terms.php?ref=pf>

La propiedad es del usuario, pero se cede el uso sin control y la obtención de beneficios a

Facebook sin contraprestación. Twitter también muerde la manzana envenenada, parece que incluso con más ahínco: “El usuario se reserva los derechos de cualquier contenido enviado, publicado o presentado a través de los Servicios [...], el usuario otorga a Twitter licencia mundial, no exclusiva, libre de regalías (con derecho a la concesión de la licencia a terceros) para utilizar, copiar, reproducir, procesar, adaptar, modificar, publicar, transmitir, mostrar y distribuir dicho contenido en cualquier medio de comunicación o método de distribución (actual o futuro) [...]. Twitter, otras compañías, organizaciones o personas asociadas con Twitter podrán llevar a cabo tales usos adicionales sin compensar de ninguna manera al usuario que ha enviado, publicado, transmitido o puesto a disposición contenido a través de los Servicios”.

<http://twitter.com/tos>

“El dueño de los datos de un blog es su autor, no la empresa que instaló un CMS y cobra por su mantenimiento”

4. Hay que dejar claro de quién son los datos

Esta avaricia de datos no es exclusiva de los servicios de la web 2.0. En los dos últimos años hemos visto cómo el uso de los silos de datos ha sido motivo de fricción entre proveedores y usuarios del mundo de la información. En 2009 Zotero fue amenazado judicialmente por Thomson Reuters, que intentaba evitar que los usuarios exportasen e importasen información entre las diferentes aplicaciones, llegando incluso a intentar atribuirse la propiedad intelectual de los ficheros de estilos que creaban los usuarios.

<http://tramullas.com/2009/08/15/zotero-gana-una-batalla>

Se trataba de un movimiento para intentar controlar los silos personales de datos bibliográficos, dado que su gestor bibliográfico de referencias estaba siendo superado por otros más avanzados, en un movimiento defensivo evidentemente condenado al fracaso, como han demostrado posteriormente CiteULike, 2collab o Mendeley.

Las bibliotecas tampoco han quedado al margen de estas tensiones. Ocllc intentó evitar que Skyriver (que le estaba ganando contratos de servicios) utilizase los datos contenidos en WorldCat. Para ello cambió unilateralmente los términos de uso de su producto sin avisar, y llegó a enfrentarse



incluso a sus propios socios. El asunto llegó a los tribunales en julio de 2010.

<http://www.librarytechnology.org/web/breeding/skyriver-vs-oclc/>

Los problemas del catálogo colectivo C17, cuyo acceso es financiado todavía por el Instituto de Salud Carlos III, son un indicador de la inocencia con la que se han ido firmando contratos en los cuales no se había previsto la propiedad y explotación de los silos de datos creados con las aplicaciones. La web del Catálogo colectivo de Rebiun no contiene ni una mención de la propiedad de los datos del catálogo, ni de cuáles son los usos permitidos de los mismos a terceras partes. En el caso de Rebeca, se establece la necesidad de solicitar una cuenta de usuario, al parecer restringida a bibliotecas públicas, pero tampoco se ofrece mayor información. Más curioso es el caso de Dialnet, que específicamente indica, al pie de su interfaz, “todos los derechos reservados”, sin mayor aclaración, pero que en sus páginas de ayuda publica su política de acceso abierto.

http://dialnet.unirioja.es/ayuda/SPA/ayuda_oai.jsp

Todas estas cuestiones nos deberían llevar a una reflexión sobre cuáles han de ser las políticas de propiedad y uso de los datos generados por las unidades de información en el marco de su trabajo diario. Para mí evidentemente mis datos son míos y la compra de licencias de uso de una aplicación dada no debe suponer que los datos sean cedidos al proveedor, ni que tenga derechos sobre los mismos, y menos cuando son fruto de mi trabajo y además pueden ser resultado de actividades financiadas con fondos públicos.

La Ley de propiedad intelectual contempla en su artículo 12.2., una definición “amplia” de lo que se considera base de datos, y añade en el 12.3: “La protección reconocida a las bases de datos en virtud del presente artículo no se aplicará a los programas de ordenador utilizados en la fabricación o en el funcionamiento de bases de datos accesibles por medios electrónicos”,

reconociendo que no es lo mismo la base de datos (el contenido) que la aplicación usada para gestionarla.

El artículo 133.3a establece como "fabricante" de la base de datos –y poseedor de sus derechos–, a "la persona natural o jurídica que toma la iniciativa y asume el riesgo de efectuar las inversiones sustanciales orientadas a la obtención, verificación o presentación de su contenido".

El dueño de los datos de un blog o de de una web dinámica es su autor o creador de contenidos, no la empresa que instaló un CMS y cobra por su mantenimiento. Por si acaso, que no se olvide

hacerlo constar de manera expresa en el contrato correspondiente.

En realidad las administraciones públicas españolas ya están entendiendo la importancia de los silos de datos abiertos en cuanto han empezado a dar acceso a información mediante iniciativas de apertura de datos (*Open data*) para que puedan ser usados por terceros. Será interesante ver cómo se van conjugando las iniciativas privadas en busca de rentabilidad con la utilización sensata de los datos de las administraciones públicas que, no debe olvidarse, se sufragan con el dinero de los ciudadanos.

EPI y ThinkEPI en Fesabid

25-27 mayo 2011

Ajustada al lema de las 12JED "Una profesión, un futuro", así como a 2 de sus ejes temáticos: 1.Una profesión que evoluciona y 2.Una profesión que innova, el Grupo ThinkEPI ha organizado el panel "**Gestionar el mañana**".

Intervienen:

Tendencias en el diseño de webs

Enrique Orduña, UPV

Criterios para innovar en nuestro trabajo

Pablo Lara-Navarra, UOC

Novedades en móviles. Qué hay de realidad práctica.

Natalia Arroyo, FGSR

Tecnologías emergentes a las que hay que prestar atención

Jorge Serrano-Cobos, MASmedios

Nuevas salidas profesionales

Tomàs Baiget, EPI

Moderador: **Javier Guallar**, UB, UOC y URL

Miércoles **25 de mayo**, de 12.30-14h

Auditorio del Palacio de Ferias y Congresos de Málaga

<http://www.fesabid.org/malaga2011/programa-cientifico>

EPI tendrá un stand en la Feria Documat