

data404: producir, liberar, remezclar y representar la información pública

Como la mayoría de proyectos, **data404** comenzó como una idea vaga sobre varios tópicos en los que sólo tenían en común el concepto de abierto y libre, pero tampoco tenía un claro objetivo de lo que pretendía ser; nació en gran medida por la frustración al comprobar como las administraciones públicas actuaban como cajas negras cuando se trataba de acceder a los datos. Cory Doctorow¹ tiene toda la razón al extender el principio general de la revista Make² al sector público, *If you can't open government, you don't own it*. Los datos públicos deberían, por su propia condición, ser accesibles y libres a todo tipo de restricción legal que impida su análisis, distribución, reutilización, remezcla y representación. De esta manera, la visualización de datos públicos se entendería no sólo como actividad puramente estética, sino como estrategia para contar una historia en la que los protagonistas de la narración serían precisamente los datos y, la trama de la historia correspondería a revelar una información que, hasta el momento, permanece oculta, invisible en la uniformidad de los datos en crudo.

data404 es, por el momento, un recolector³ de noticias, artículos, opiniones, proyectos, documentación en torno a dos movimientos que convergen, sin problemas, en un frente común, el *open data* y el *open government*, y que pueden situarse en un contexto más amplio al que el teórico de [Peer-to-peer](#) Michel Bauwens⁴ da a conocer como *Openness*. Los dos movimientos, en conjunto, pueden resumirse como el intento de hacer un gobierno transparente y próximo a la ciudadanía si aquel libera la totalidad de los datos que produce de forma transparente. Si el sector público siguiera este principio de transparencia, tacharía a gusto y sin mayor objeción los verbos *producir* y *liberar* de la comunicación, centrándose en el lado más creativo de la manipulación de datos, la remezcla y la visualización. Pero como todavía no vivimos en ese escenario ideal, y los datos no están disponibles para que pueda reutilizarlos, remezclarlos y representarlos, no tengo otra opción que pensar en términos de liberación de datos -o crearlos cuando estos no existen-, siguiendo los pasos del *frente de liberación de datos* creado por un reducido grupo de ingenieros en el seno de Google:

Users should be able to control the data they store in any of Google's products. Our team's goal is to make it easier to move data in and out.⁵



No es difícil trasladar este sencillo mensaje a la realidad de la ciudadanía. Los ciudadanos deberíamos tener el control de los datos que produce y gestiona cualquiera de las diferentes administraciones y organizaciones del sector público. El objetivo no sería otro que facilitar todas las herramientas y técnicas libres, disponibles en la actualidad, que hagan posible la liberación de los datos del sector público.

data404 es también un juego muy simple de palabras que atribuye el significado *slang* del número 404⁶ a aquellos datos que están disponibles pero que ni siquiera encontramos con los buscadores principales como Google o Bing. Estos datos acostumbra a estar detrás de un oscuro y barroco formulario y, por lo tanto, sin otra alternativa que realizar peticiones una y otra vez si deseamos obtener la información que precisamos. Esta y toda información que permanece detrás de un formulario web de consulta pertenece a la vasta región de la web profunda⁷ (*deep web*) y presenta la particularidad que es una información muy difícil o imposible de indexar por los buscadores. Otra ocurrencia 404 del sector público es proporcionar una extensa lista de datos, en estructura tabular, encerrada en un formato final -como por ejemplo, en PDF- lo que impide que nadie pueda reutilizar la información que allí está atrapada no sin grandes dificultades.

Veamos un ejemplo. La lista de centros públicos y concertados de Catalunya sólo está disponible, hasta el momento, detrás del siguiente formulario http://www10.gencat.net/pls/ense_ensenyam/p01.menu del portal de la Generalitat de Catalunya. Este mismo formulario se repite de nuevo en otro servicio al ciudadano de la Generalitat: <http://tr.im/Ez01> (me he visto obligado a reducir la inacabable URL original del servicio). Si bien es cierto que el formulario me permite acceder a una información *precisa*, eligiendo las opciones adecuadas de aquel en relación al objeto de búsqueda, el servicio no es neutral desde el momento que el acceso a la lista de centros está gestionada y organizada por la propia infraestructura web. Para colmo, el resultado es una página web en la que ni siquiera la información del centro está correctamente estructurada, lo que dificulta todavía más la transportabilidad o interoperabilidad de los datos que constituyen la ficha del centro educativo a otras herramientas o infraestructuras web. En definitiva, proporcionan un documento fragmentado destinado a ser únicamente de lectura para humanos, pero impracticable o inútil para cualquier otro *software*.

Este escenario no es casual. En realidad, el sector público, en su expresión digital, es un importante enclave de la *deep web* cuando está distribuyendo los datos detrás de barrocos formularios, o cautivos en formatos finales del tipo PDF.

Beth Simone, en su libro *Wiki Government. How technology can make government better, democracy stronger*⁸, defiende la idea que los datos públicos han de ser transparentes si pueden ser reutilizados en beneficio propio de la ciudadanía. Este principio de transparencia de los datos públicos no es posible si falla uno de las tres condiciones siguientes:

- Todos los datos públicos han de ser **accesibles**.
- Todos los datos públicos han de ser **localizables**.
- Todos los datos públicos han de ser **usables**.

Datos accesibles

El grupo de trabajo sobre Open Government, [Got Data!](#), de los que forman parte Lawrence Lessig, Tim O'Reilly o Tom Steinberg (de [Mysociety](#)), establecieron ocho principios que una administración abierta debería cumplir con el objeto de ser más eficiente, transparente y relevante para la ciudadanía. Uno de ellos se concretó en la accesibilidad de los datos, también entendido como *Data are available to the widest range of users for the widest range of purposes*, es decir, la accesibilidad debe cubrir todo el espectro de usuarios finales como el amplio abanico de posibilidades que nos permita las tecnologías en la actualidad. También es oportuno añadir que la accesibilidad de los datos públicos no podrá serlo si su publicación y distribución en la red no está bajo licencias no restrictivas o encerradas en cajas negras en forma de formatos finales o formularios. Si los datos están ubicados detrás de una interface interactiva, se ha de proporcionar el bloque completo de datos en bruto y en diferentes tipos de formatos abiertos acompañados de la suficiente documentación para que el destinatario comprenda su naturaleza. Hace unos meses, Alberto Ortiz de Zárate, más conocido en la red como *alorza*, pedía en el seno del blog [Administraciones en red](#) la colaboración en definir el proyecto de apertura de datos del Gobierno Vasco⁹.

La respuesta de un lector manifestó claramente la necesidad de disponer los datos del sector público:

¿Por qué debería haber limitaciones más allá de las lógicas derivadas de la LOPD? Salvando estas cuestiones, cualquier dato "bruto" generado por el sector público es de la sociedad y debería ponerse a su disposición. Nunca he entendido ese cerramiento de mollera de que lo que produce la Administración es de la Administración y datos sobre actividad es desde luego lo que más hay. ocortes — 29-07-2009 - 17:00:17 GMT 2

Datos localizables

El sector público ha de trabajar con infraestructuras que faciliten la navegación con el fin de localizar sin dificultad los datos que necesitamos. No podemos delegar esta competencia a los buscadores generales ya que, en la mayoría de casos, los datos del sector público no pueden ser indexados por diferentes motivos (contenidos públicos detrás de servicios que requieran autenticación, formularios, formatos finales, información no enlazada...). La oportunidad de localizar datos públicos no acaba en los límites del habitual mapa de la web. Amber Simmons, en su artículo [You Can Get There From Here: Websites for Learners](#) defiende, al margen de crear páginas web que sigan los principios de la W3C, que separen la estructura de la página del diseño, la incorporación de elementos narrativos que inviten a los usuarios a una exploración *irregular* de los contenidos. Estimular la serendepidad en lugar de trazar rutas lineales de acceso a la información. Marc Augé en su *Elogio de la bicicleta*¹⁰ describe una experiencia similar de aquellos que cogen por primera vez una bicicleta en la ciudad, si la entendemos como metáfora de la navegación por las infraestructuras del sector público.

les permite reevaluar las distancias y hacer acercamientos que les están vedados en el transporte público, sujeto a itinerarios fijos.

Ejemplos como la Wikipedia o Delicious, donde el usuario navega por una secuencia de enlaces que permite la exploración y el encuentro fortuito de nueva información cuya existencia no conocía anteriormente.

Datos usables

¿Qué entendemos por usabilidad de los datos públicos? En palabras de Beth Simone, la usabilidad es una cuestión que atañe al consumo de los datos del sector público. A diferencia de un consumo pasivo y unidireccional de los datos, se trata de ofrecer los datos en formatos que promuevan su reutilización., que estimulen el debate y el análisis. El formato de los datos ha de asegurar, por lo tanto, la lectura tanto por humanos como por máquinas sin exclusión social o legal, bajo términos de seguridad y autenticidad. La usabilidad de los datos públicos sitúa el acento en la participación de la ciudadanía en la construcción de la información pública y desplaza, definitivamente, cualquier análisis, estudio o informe de los datos públicos que surja de la propia administración.

La reutilización de la información del sector público consiste en el uso por parte de personas físicas o jurídicas, de información generada por organismos del sector público, con fines comerciales o no comerciales. No obstante, la Ley 37/2007 **excluye** del concepto de reutilización **el intercambio de información entre Administraciones y organismos del sector público en el ejercicio de las funciones públicas que tengan atribuidas.** [Proyecto Aporta](#)

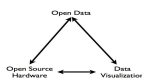
data404 como *help framework*

¿Qué hacer cuando este principio de transparencia no está presente en el sector público? La lista de centros educativos de Catalunya es un ejemplo de falta de transparencia. Los datos de las fichas de centros no son accesibles, localizables y usables. Como usuario no encuentro ninguna alternativa en forma

de archivo en un formato abierto que proporcione la lista completa de centros. Es un ejemplo evidente de 404: los datos están disponibles bajo la estructura de la administración luego, la oportunidad como ciudadano de reutilizar, analizar, estudiar, remezclar o cruzar estos datos públicos con otros (índices de fracaso escolar, paro, inmigración, estabilidad de la plantilla del profesorado, número de servicios culturales, etc.) es nula.

Un escenario modelo como el anterior, en el que el sector público no participa de una versión abierta de los datos públicos; el marco conceptual de abertura que describe el mapa mental de Michel Bauwens, [Everything Open and Free](#); la aparición de iniciativas como [CivicDB](#)¹¹, [ReWired State](#)¹², [Fix my Street](#)¹³ o la [Sunlight Foundation](#)¹⁴ y, finalmente, la lectura del artículo *Visualizing Urban Data*¹⁵, de Michal Migurski, en el que explica la historia de la creación y desarrollo del servicio [Crimespotting](#) en respuesta al deficiente servicio oficial del departamento de policía de la ciudad de San Francisco, han servido de gran ayuda en la definición de un proyecto embrionario como es ahora mismo **data404**.

En primer lugar, **data404** es sensible a los tres vértices que dibujan el siguiente triángulo:



1. **Open data**, tanto desde un punto de vista técnico como naïf: la obertura de datos del sector público de la administración pública como la generación y producción de datos por parte de la ciudadanía ([Fix my Street](#) es un buen ejemplo) pueden hacer de este mundo un sitio mejor.
2. **Open source hardware**, o lo que es lo mismo, proyectos como [Arduino](#), [BugLabs](#) o la placa [Beagle](#), la revista [Make](#), en el que se estimula la creación y liberación de todos los ingredientes que conforman la totalidad del proceso de producción de un dispositivo, va cristalizando en un enorme interés colectivo. Crear y diseñar pequeños y medianos dispositivos en la actualidad ya no resulta ni tan caro ni complejo. Dispositivos que actuarían, en este contexto, como sensores y, por lo tanto, proporcionarían a la ciudadanía la capacidad de convertirse en generadores de datos.
3. La disciplina de la **visualización de datos** o la oportunidad como ciudadano de *opinar* (desde la reflexión a la crítica) desde una perspectiva visual, no en términos estéticos y sí en clave de comunicación. Revelar nuevos significados a partir de la visualización de datos que, de otra manera, en su representación más cruda, permanecerían ocultos. De acuerdo con Edward Tufte, la excelencia de la representación gráfica de datos consiste en que ideas complejas se comuniquen con claridad, precisión y eficiencia.

data404 pretende recoger, agrupar y ordenar toda documentación que proporcione un marco de ayuda conceptual y técnica para quienes deseen emprender una acción de liberar, remezclar, producir y representar datos del sector público. **data404** le gustaría ser, en definitiva, una suerte de *help framework* -siguiendo el viejo patrón MVC (*Model*, *View* y *Controller*)- en el que la documentación estaría agrupada en

- **Models**: concentrar toda la documentación y ejemplos prácticos sobre herramientas que permitan la liberación de datos. Liberar adquiere aquí el significado de convertir un conjunto de datos públicos, cautivos detrás de formatos finales o ofuscados formularios, en formatos abiertos del tipo XML, JSON, CSV o KML. Se trataría de dar a conocer las técnicas conocidas como *web scraping* o *screen scraping* gracias al uso de librerías como [scrubyt](#) o [beautifulsoup](#)) a herramientas más sofisticadas del tipo [Yahoo Pipes](#), [Dapper](#) o incluso la hoja de cálculo de Google Docs.
- **Controllers**: este apartado estaría reservado a dar a conocer estrategias de creación ágil y flexible de *web services* y *mashups* de los datos públicos liberados gracias al uso de *microframeworks web* actuales como [Sinatra](#) o [WebPy](#). Estos micro-servicios podrían dar una respuesta idónea a la necesidad de remezclar la información al constituirse aquéllos como extractores y proveedores de datos públicos en bruto. En esta sección también se incluiría el aprendizaje y formación práctica de

proyectos de *open source hardware* que contribuyan a la creación de pequeños dispositivos que actúen como sensores y, por lo tanto, sean capaces de adquirir y recoger datos del exterior, procesarlos y distribuirlos a través de *web services*, mencionados más arriba en este mismo apartado.

- **Views:** el último componente estaría orientado a cómo utilizar las diferentes técnicas de visualización en la transformación mera información en conocimiento. Aquí incluyo el interés en en la democratización de la visualización. No puedo estar más de acuerdo con la visión de Jon Udell en su entrada [Visualization Trends For The Noosphere](#) y que resume perfectamente en la siguiente observación:

To democratize visualization we'll need a new generation of software. With personal and then web-based computing, we've seen it happen again and again: spreadsheets, desktop publishing, web multimedia, cloud-based services. Now, across a range of devices as well as in the cloud, we have the raw technologies to democratize the visualization -- and collaborative analysis -- of data.

Estas herramientas ya están disponibles en la actualidad. Pueden ser de propósito general, provenientes de la creación gráfica procedural como es el caso del enormemente popular Processing¹⁶ o o Nodebox¹⁷, pero también aplicaciones web orientadas al usuario final como Many Eyes¹⁸, Swivel¹⁹, Data360²⁰. Pero también merece una enorme atención el goteo continuo de *frameworks*, *toolkits* o librerías realizadas en lenguaje javascript, como por ejemplo Raphaël²¹, que permiten una mayor presencia de la práctica de la visualización de datos en la web.

Por último, **data404** persigue un objetivo claro que creo común entre los diferentes proyectos y comunicaciones de Visualizar'09: difundir y estimular la importancia del acceso, remezcla y visualización de los datos de las administraciones públicas en beneficio de la comunidad. Este escenario ideal supondría la migración de una ciudadanía que abandonaría el papel de consumidor pasivo de la información del sector público a un rol activo y creador como agente transformador de datos en bruto en un valioso conocimiento colectivo.

Xavier Belanche Alonso, Barcelona, 10 de noviembre de 2009

-
1. [if you can't open it, you don't own it](#) ↵
 2. [If You Can't Open Government, You Don't Own It](#), Makezine Magazine, Vol. 18, pág. 17 ↵
 3. <http://www.twine.com/twine/126zm5sh0-1q3/data404> ↵
 4. [Open Everything Mindmap and visualization] (<http://blog.p2pfoundation.net/open-everything-mindmap-and-visualization/2009/09/08>) ↵
 5. [The Data Liberation Front](#) ↵
 6. El número 404 tiene un significado familiar para todos los usuarios de la red. Cada vez que realizamos una petición de búsqueda de una página web que no existe, el servidor nos devuelve un mensaje de error clasificado como *404*. Como ya viene siendo habitual, el significado técnico 404 ha sido adoptado por la jerga de calle, especialmente en el área de Londres ([Don't be 404, know the tech slang](#)): en *slang*, 404 es sinónimo de *ni idea*. ↵
 7. http://en.wikipedia.org/wiki/Deep_Web ↵
 8. Simone, Beth, *Wiki Government. How technology can make government better, democracy stronger*, (2009) pág. 121. ↵

9. <http://eadminblog.net/post/2009/07/27/solicito-ayuda-para-definir-el-proyecto-de-apertura-de-datos-en-gobierno-vasco> ↩
10. Augé, Marc, *Elogio de la bicicleta* (2009), pàg. 65-66. ↩
11. *CivicDB helps government organizations liberate data in multiple open formats like CSV, XML, RESTful web services by providing a full end-to-end solution.* ↩
12. *Government isn't very good at computers.They spend millions to produce mediocre websites, hide away really useful public information and generally get it wrong. Which is a shame.* ↩
13. *FixMyStreet is a site to help people report, view, or discuss local problems they've found to their local council by simply locating them on a map. It launched in early February 2007.* ↩
14. *Turning Government Data Into Usable Information idea is born government data is a mess developer takes messy data and converts it into user friendly code our products are put back out for the public to see and use* ↩
15. Migurski, Michal, *Visualizing Urban Data*, (2009), en *Beautiful Data. The Stories Behind Elegant Data Solutions*, O'Reilly. ↩
16. <http://processing.org/> ↩
17. <http://nodebox.net/code/index.php/Home> ↩
18. <http://manyeyes.alphaworks.ibm.com/manyeyes/> ↩
19. <http://www.swivel.com/> ↩
20. Yau, Nathan (2009) *The Flowing Data guide to visualisations* <http://www.guardian.co.uk/news/datablog/2009/jun/15/google-ibm> ↩
21. <http://raphaeljs.com/> ↩