



Big Data y los nuevos métodos de visualización de la información

En el contexto actual, es necesario plantear nuevas formas de organizar y presentar la información al usuario, sea este un científico que trabaja con los enormes resultados de un acelerador de partículas o un ciudadano normal que se quiere informar sobre el histórico de datos del clima de su ciudad obtenidos de un gran número de sensores ambientales.

La expresión *Big Data* se refiere a grandes conjuntos de datos digitales que requieren de sistemas informáticos para su captura, almacenamiento, búsqueda, manipulación y visualización. El aumento del número de fuentes y la potenciación de la captura implícita y explícita de datos a través de *smartphones*, *tablets* o multitud de nuevos sensores digitales han planteado una nueva situación en lo que al tratamiento de los datos se refiere.

La revolución digital ha transformado nuestra forma de relacionarnos con el entorno y de construir el conocimiento. No estamos ante un cambio de paradigma tradicional en el que es posible definir unas fases de entrada en crisis, revolución e instauración del nuevo paradigma. El esquema clásico de Kuhn¹ no sirve cuando hablamos de tecnología y no nos limitamos a juzgar un caso concreto. La sociedad tecnológica es dinámica y cambiante y rehúye, implacable, las teorías sistémicas con las que nos hemos acostumbrado a construir el conocimiento. La variable ha tomado un papel protagonista en la sociedad actual y ha reducido la vigencia de las reflexiones, cuando tratan de tecnología, hasta puntos en los que, más que resultados, el mundo académico propone caminos a seguir.

Uno de los cambios más importantes derivado directamente de las nuevas tecnologías es el aumento de nuestra capacidad de generar, almacenar y gestionar datos. La ampliación constante de la capacidad de los ordenadores y los dispositivos de almacenamiento, la participación efectiva del usuario convertido en *prosumer* (consumidor proactivo) y el aumento del volumen de sensores activos y pasivos que recopilan información han dado un vuelco a cualquier ecuación que pretenda definir el crecimiento del corpus del conocimiento.

1. Kuhn, T. *The Structure of Scientific Revolution*. Chicago: University of Chicago Press, 1962.

Si en 2008 aumentamos el volumen de datos en 487 billones de gigabytes (GB), se calcula que en 2012 esta cantidad se multiplicó por cinco, y no hay ninguna tendencia prevista para la reducción de este crecimiento. Por lo tanto, el reto de mantener y transmitir el conocimiento se encuentra tan solo en su primera fase. Hemos alcanzado un zettabyte (ZB) de almacenamiento digital, el equivalente a más de mil billones de los disquetes de hace unos años.²

Los retos

El filósofo David Casacuberta planteaba en su blog que, si bien somos —sin ninguna duda— las generación que más información han generado (valga la redundancia), podemos ser también las primeras en no ser capaces de transmitirla. Y esto es así porque, como en la mayoría de fenómenos tecnosociales, la reflexión ética y filosófica —incluso en el seno de las disciplinas tecnológicas— no puede avanzar a la misma velocidad que las nuevas tecnologías y sus implicaciones.

“Y aquí tenemos la gran paradoja de nuestra cultura: tenemos el mayor registro —con diferencia— de acciones culturales, pero está en formatos que no resistirán el paso del tiempo. ¿Qué quedará realmente de las campañas *online* de Barack Obama? ¿Dónde está ahora la información amorosamente guardada en las páginas *vintage* de Geocities? ¿Quedan todavía copias del Wordstar? ¿Se pueden ejecutar todavía? ¿Cuánta información se ha perdido para siempre cuando una institución decide llenar un *container* de discos *floppy* inservibles y tirarlos a la basura?”³

El reto de gestionar y transmitir el conocimiento adquirido es, seguramente, uno de los mayores que los seres humanos tenemos por delante como especie social; y la única manera de acometerlo es dedicar tiempo y recursos a avanzar en la gestión y la clasificación de los datos, desarrollando nuevos métodos para visualizar la información. La gran potencialidad que ofrecen los sistemas de procesamiento del *Big Data* como base para el análisis cultural parece ser una de las posibles soluciones para transformar los datos en conclusiones y poderlos “archivar”. Esto significa que, si bien es muy difícil que podamos transmitir la totalidad de los mensajes de Facebook que constituyeron la campaña en línea de Barack Obama, sí que podemos transformar estos datos en infografías, análisis o recopilaciones y recoger en ellos su esencia útil que, de lo contrario, perderíamos. Solo de este modo garantizaremos una continuidad en la historia del conocimiento y permitiremos a las generaciones futuras seguir estudiando el pasado y seguir aprendiendo de él.

No es posible hacer copias de seguridad de las relaciones entre las cosas, no es posible subir a la Nube la interacción entre una sociedad y los *trending topics* de Twitter. No podemos archivar la naturaleza interconectada de los datos, por lo que el análisis cultural derivado de ellos es aún mucho más importante.

Los datos son hoy en día la base del conocimiento; sin ellos no hay tecnología (ni medicina ni combustibles alternativos, por ejemplo). Y, sin el análisis de estos datos, no se pueden hacer radiografías sociales, culturales ni económicas serias. Sin una reflexión profunda sobre sus usos, sus limitaciones y los derechos éticos

2 Sáez, C. *¿Dónde están mis datos?* [en línea]. [Consulta: 8 de marzo de 2013]. Disponible en: <http://www.lavanguardia.com/estilos-de-vida/20120719/54326650737/donde-estan-mis-datos.html>

3 Casacuberta, D. *Entender nuestra cultura, net art y el Big Data* [en línea]. [Consulta: 8 de abril de 2013]. Disponible en: <http://elbigdata.net/entender-nuestra-cultura-net-art-y-el-big-data/>

relacionados, los datos solo son datos. Es frecuente la confusión —incluso en muchos artículos académicos— entre “datos” e “información”, confusión que se da en la aplicación de los conceptos y no en su definición. Los datos no son útiles por sí solos: deben asociarse a algo para ofrecer un significado, un conocimiento, una conclusión. No constituyen información ni tienen capacidad de comunicar un significado. El conjunto organizado de datos procesados se convierte en información, y es entonces cuando adquieren la utilidad que como datos independientes no tenían.

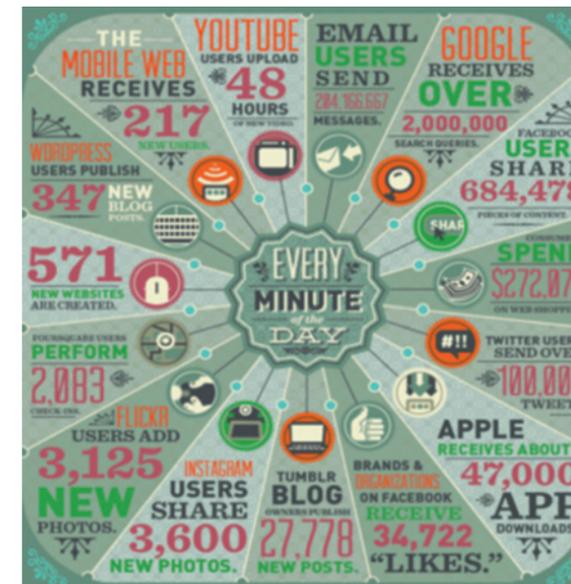
Por lo tanto, debemos afrontar tareas importantes en la gestión de la información, entendiéndose por *gestión* no solo el almacenamiento y la recuperación, sino también el proceso de transformación de los datos y la visualización de la información. Y esto es válido tanto en entornos científicos y académicos como en cualquier medio social. Los datos son el elemento sobre el que hemos aprendido a construir el conocimiento científico y, desde hace algunas décadas, también el conocimiento social. Lo que no está tan claro es que estemos preparados para sacar partido a los datos cuando estos se miden en miles de millones de *bytes*.

La cultura del *Big Data*

En un contexto en el que los avances tecnológicos dejan de ir de la mano de la reflexión académica con mezquina facilidad, es necesario hacer un *stop and go* profundo y plantearnos cuál será la senda que nos llevará a la integración del *Big Data* como un elemento central en las investigaciones contemporáneas, y de qué forma ofreceremos al público el resultado de las mismas.

La sociedad contemporánea está regida por los grandes poderes económicos. En el caso del *Big Data* son las grandes corporaciones las que se han encargado de iniciar la carrera comercial en torno a los datos, para asegurarse una buena porción de un pastel llamado a ser el más goloso de los próximos años. Así, las empresas, entidades o gobiernos que están destinando recursos a investigar y desarrollar las potencialidades del *Big Data* serán las que recogerán —en un futuro muy cercano— los frutos de la gran cosecha de los datos. Sin embargo, mientras que el *Big Data* y sus implicaciones ocupan un gran número de másteres y postgrados en las mejores universidades norteamericanas, y en reconocidos centros de investigación como el MIT, en España aún es muy difícil atraer la atención del mundo académico y promover nuevas iniciativas al respecto.

El *Big Data*, más allá de dar pie a nuevas y prometedoras ocupaciones para científicos y analistas de datos,⁴ es un fenómeno cultural que nos afecta a todos.



▲ DOMO. *Every Minute of the Day*, 2012. (Fuente: <http://img.xatakaciencia.com/2012/07/informacion-internet-minuto.png>).

4 Davenport, T.; Patil, D. J. *Data Scientist: The Sexiest Job of the 21st Century* [en línea]. Harvard: Harvard Business Review [Consulta: 8 de abril de 2013]. Disponible en: <http://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/>

y descubrir nuevas posiciones, facetas o enfoques para avanzar en su resolución. Y, por último, una buena visualización de los datos contribuye a crear puntos de vista compartidos de una situación concreta y a agrupar a los expertos en torno a las acciones necesarias.

Los nuevos métodos para tratar la información se presentan como garantía de una correcta interpretación de los datos y de su aprovechamiento al servicio de la creación de conocimiento. Visualizando podemos contribuir a confirmar o refutar una teoría social o política, podemos mostrar algo que ya sabemos —narrando los hechos— y podemos también abrir nuevas vías de exploración transformando grandes conjuntos de datos —muy difíciles de gestionar y analizar— en modelos que nos permitan predecir el comportamiento de un sistema y aprender a manejarlo mejor.

“Creo que migraremos con gran naturalidad hacia visualizaciones superiores que nos ayudarán a sobrevivir en este mar de información. Desde los tiempos de las pinturas rupestres, la representación gráfica siempre ha sido una parte integral de la forma en que las personas piensan, se comunican y entienden el mundo. En el mundo actual, los nuevos sistemas de información son centrales en cualquier proceso de gestión y actividad organizativa.”⁸

Un interesante ejemplo de aplicación de visualización de información para el desarrollo de la ciencia es el proyecto *Planet Hunters*, diseñado y desarrollado por el estudio madrileño Vizzuality para la NASA con la colaboración de equipos de la University of Oxford. Planethunters.org es una plataforma abierta a astrónomos profesionales y aficionados, cuyo objetivo final es detectar planetas fuera de nuestro sistema solar.

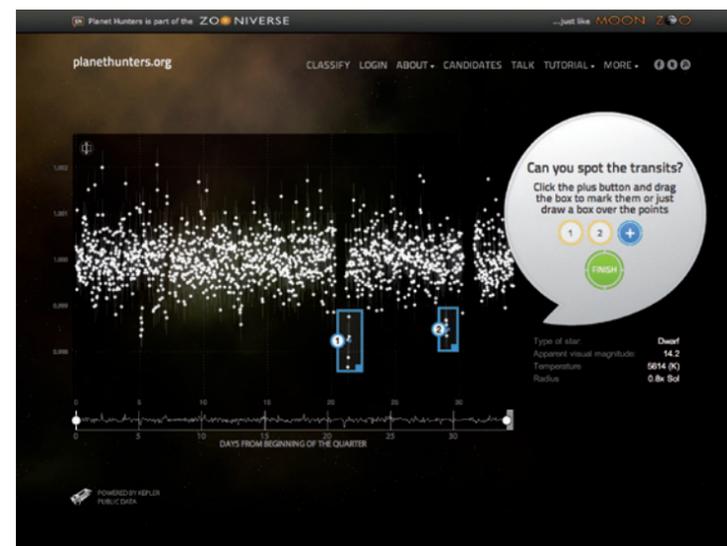
Planet Hunters pretende visualizar la ingente cantidad de datos provenientes de la sonda Kepler, que se encuentra en órbita alrededor del Sol. Si bien la NASA utiliza poderosísimos ordenadores para analizar estos datos, en este proyecto se busca una presentación visualmente accesible de los mismos para que sean inteligibles para la mente humana, aprovechando la gran capacidad que tenemos para el reconocimiento de patrones visuales. Los datos visualizados son medidas de luminosidad de 150.000 estrellas tomadas cada treinta minutos. Los usuarios de la plataforma intentan detectar los cambios de luminosidad, ya que pueden ser indicadores de la existencia de un planeta. Los seres humanos ven cosas que los algoritmos de los ordenadores más potentes no son capaces de detectar, como en el caso de planetas pequeños del tamaño de la Tierra.

Data scientists y data designers

Sin embargo, un análisis del *Big Data* y de sus implicaciones para la visualización de la información no estaría completo si no se explicara qué herramientas o conocimientos vamos a necesitar para la aplicación de estos nuevos métodos. Amanda Cox, editora gráfica de *The New York Times*, afirma que las cualidades necesarias para su desarrollo son ya muy difíciles de encontrar en un solo profesional. La programación, la estadística y el diseño gráfico se aúnan en un trabajo que debe ser colaborativo, suma de distintas visiones y capacidades.⁹ Tras todo

8 *Ibidem*.

9 Berinatto, S. *The Power of Visualization's "Aha!" Moments* [en línea]. [Consulta: 8 de abril de 2013]. Disponible en: http://blogs.hbr.org/hbr/hbreditors/2013/03/power_of_visualizations_aha_moment.html



▲ Vizzuality. Interfaz del proyecto *Planet Hunters*, 2011. (Fuente: www.vizzuality.com)

10 Manovich, L. *Software Studies: Cultural Analytics* [en línea]. [Consulta: 8 de abril de 2013]. Disponible en: <http://lab.softwarestudies.com/2008/09/cultural-analytics.html#1>

11 *Ibidem*.

buen proyecto de visualización debe haber un buen diseño, un código efectivo y una buena gestión y análisis de los datos.

Solo así es posible llegar a buen puerto, estableciendo una marcada jerarquía informacional y promoviendo el estudio académico sobre cómo percibimos la información, cómo la gestionamos y cuáles son nuestras intuiciones al respecto. Así pondremos el código informático al servicio de un nuevo objetivo cultural. Para transformar datos en información, debemos priorizar ese objetivo informacional por encima de la estadística oscura y poco funcional.

Si bien el *Big Data* centra actualmente todos los debates y es difícil determinar cuáles son las iniciativas que van a conseguir realmente avances significativos, hay algunos proyectos que en estos momentos ya han hecho grandes aportes. Manovich —uno de los grandes teóricos de los nuevos medios— se dio cuenta hace tiempo de la necesidad de poner el *software* al servicio del análisis cultural, e inició un

gran proyecto en torno a lo que él mismo llamó *Cultural Analytics*,¹⁰ que no es más que el proceso de estructurar la información y los métodos de recuperación y visualización con vistas a la eficacia de dicho análisis cultural.

“*Cultural Analytics* es el uso de métodos computacionales y de visualización para el análisis de grandes conjuntos y flujos de datos culturales. [...] Podemos llegar a una comprensión más completa e inclusiva de la evolución y la dinámica culturales humanas recurriendo a todos los artefactos culturales generados ya digitalmente o digitalizados *a posteriori* en cualquier medio de toda la historia de la humanidad. Podemos ubicar en el mapa ‘en tiempo real’ los patrones culturales que van cambiando por todo el mundo, utilizando todos los datos disponibles en la Red. Alcanzar estas metas pasará no solo por resolver numerosos problemas técnicos sino también por desarrollar nuevas metodologías y conceptos para el análisis cultural basado en los datos.”¹¹

Para desarrollar el proyecto, Manovich se dio cuenta de la necesidad de programar nuevas herramientas informáticas que le permitieran digitalizar grandes cantidades de información y procesarla para obtener mapas visuales realmente útiles. Como comentábamos, la única vía para una visualización de datos eficaz es la suma de los conocimientos de diseñadores, programadores y analistas en pos de un objetivo común. Así pues, el perfil que debe asociarse a los retos de la información no es ya un perfil estanco entorno a una sola disciplina. Hay que perder el miedo a la tecnología digital desde el diseño y al diseño desde la tecnología digital; solo así encontraremos un espacio común de reflexión y avance que permita tratar los problemas derivados del *Big Data* con la atención y profundidad que se merecen, más allá del *boom* publicitario y centrándonos en su verdadera significación.