De los datos al Big Data

POR **SORAYA PANIAGUA**

En noviembre de 2012 tuve la gran oportunidad de organizar para Fundación Telefónica las primeras jornadas *Vivir en un mar de datos*[1]. Fueron tres días y veinticuatro ponentes. Un gran plantel de científicos, ingenieros, periodistas, ejecutivos y empresarios españoles pasaron por el escenario de la Fundación y por primera vez en España se habló de la nueva Ciencia de los Datos.

Este proyecto ha propiciado la creación del grupo Macrodatos[2], en el que también están involucrados Silvia Pradas[3], Mario Tascón[4] y Pepe de la Peña[5]. Nuestro objetivo es promover la divulgación y el conocimiento en torno al mundo de los datos.

Datos, sensores y redes de sensores

Para entender qué es *Big Data* lo primero es saber qué son los datos y dónde están.

Los datos están en todas partes; nosotros mismos, como seres humanos, generamos miles de millones de datos. Los datos están en el aire que respiramos, en la luz, en el agua, en las plantas, en antenas de telefonía, en las cámaras de vídeo o en el subsuelo de *parkings* y carreteras. Un dispositivo inteligente, como un iPhone, tiene más poder de computación del que tenía la NASA cuando puso el primer hombre sobre la Luna.

Los sensores son esas máquinas que permiten capturar las magnitudes físicas o químicas y convertirlas en datos, por ejemplo temperatura, luz, distancia, aceleración, inclinación, desplazamiento, presión, fuerza, humedad, sonido, movimiento o el pH. Los sensores existen desde hace décadas, no es nada nuevo, pero hay tres hitos que marcan un nuevo rumbo y propician la llegada de la nueva sociedad de la inteligencia: el bajo consumo, el abaratamiento de los dispositivos y las comunicaciones inalámbricas (*Wi-Fi*, bluetooth, RFID, *ZigBee*, etc.), que permiten la creación de las redes de sensores inalámbricas. Estamos asistiendo a una gran revolución en el mundo de los sensores. En España hay muy buenos ejemplos de *startups* de éxito, como Farsens[6], Libelium[7], Panstamp[8] o Cilab[9].

En un futuro no muy lejano vamos a poder vivir en ciudades inteligentes llenas de sensores, que no solo van a mejorar y optimizar nuestro estilo de vida sino también el medio ambiente.

Fundación Telefónica
Telefónica

Las redes de sensores de humedad pueden detectar peligros de incendios en bosques remotos. Los sensores de control de estrés en edificios informan sobre la integridad estructural de puentes, casas o carreteras. Una red de sensores de vibración advierte que un rodamiento está empezando a fallar. Los sensores integrados en la ropa, y en el cuerpo, proporcionan un seguimiento continuo de nuestros signos vitales. Por ejemplo pueden detectar el inicio de un ataque cardiaco o una presión arterial peligrosamente alta. Mediante diminutos sensores un diabético puede controlar los aumentos repentinos de los niveles de azúcar y automatizar un sistema de inyección de insulina.

Los sensores son, además, la base de la robótica, de la inteligencia artificial. Disciplinas que avanzan a una velocidad de vértigo. Los robots sociales y asequibles ya son una realidad, al igual que el coche que se conduce solo.

Vemos un torrente de dispositivos que van a controlar nuestro entorno de una manera que apenas podíamos imaginar hace unos años. Algunos ya están aquí. Otros están por venir.

Los datos están en los códigos de barras, en las tarjetas de crédito, en las antenas de telefonía o en los contadores de luz y agua. Las empresas y Administraciones Públicas manejan un tremendo caudal de datos, estructurados (capaces de ser tratados por máquinas) o desestructurados. El movimiento *Open Data* lo que persigue es precisamente la apertura de esos datos. Si gobiernos y empresas ponen la información a disposición de los ciudadanos se pueden crear nuevos servicios y aplicaciones que redunden en beneficio de todos. Baste ver los resultados de las políticas de Gobierno abierto[10] en Inglaterra. Otro concepto muy ligado a *Open Data* es el llamado periodismo de datos, una nueva forma de contar historias mediante increíbles visualizaciones.

Pero es quizás en Internet -y sobre todo la web social- donde se produce el mayor volumen de información. Nosotros, los ciudadanos, creamos una ingente cantidad de datos en las redes sociales. Según la UIT[11], en 2012 había dos mil trescientos millones de personas conectadas a Internet en todo el mundo. Cada minuto del día se envían más de 200 millones de *emails*, se comparten más de 700.000 piezas de contenido en Facebook, se realizan dos millones de búsquedas en Google, se envían más cien mil *tuits*, se suben 3.600 fotos a Instagram o se editan 48 horas de vídeo en YouTube. Y, cada vez más, esta actividad se realiza en movilidad. Así, un ciudadano con un *smartphone* repleto de sensores se convierte en una valiosísima fuente de datos, pero datos fuera de su control.

Hace años Google se vio en la necesidad de indexar toda la información que estaba recogiendo de Internet para poder presentar resultados significativos a sus usuarios. Google creó la tecnología subyacente de lo que hoy conocemos como *Hadoop*, una herramienta crucial que permite trabajar con miles de nodos y petabytes de datos.

Datos en poder del ciudadano

Asistimos también al asentamiento y auge del llamado *hardware* libre. Por primera vez el ciudadano, como persona individual, es capaz de crear y gestionar objetos conectados a Internet. Eso significa que los datos ya no solo están en los silos de las empresas o

BY NC ND

administraciones, están también en poder del ciudadano. Paralelamente han surgido plataformas libres donde los ciudadanos pueden volcar y compartir esos datos, plataformas como Cosm[12].

Y cuando la sociedad se enfrenta a esta nueva realidad, a este tsunami constante de datos, surgen nuevas necesidades y, por tanto, nuevas herramientas. Estamos viendo el asentamiento de lo que se denomina *Big Data,* que es la colección de tecnologías y estrategias capaces de capturar y analizar, de forma económica, grandes volúmenes de datos provenientes de múltiples fuentes heterogéneas a una alta velocidad (Josep Curto[13]). Esta definición pone sobre la mesa las famosas tres 'v' del *Big Data*: volumen, velocidad y variedad. Pero falta una cuarta variable: el tiempo real.

La parte visible del *Big Data* son las plataformas de gestión de datos. A nivel de startups, en España hay varias iniciativas que están triunfando, como Carriots[14] o Cubenube[15].

La importancia del *Big Data* radica en la capacidad de transformar grandes volúmenes de datos en inteligencia. En esa transición es donde entra en juego la llamada Ciencia de los Datos. La Ciencia de los Datos ha estado presente durante años en el mundo de la computación, pero es ahora cuando adquiere un protagonismo real y sin precedentes. Mediante complejos algoritmos, empresas como Amazon, Google, Facebook o Twitter son capaces de generar miles de recomendaciones personalizadas en tiempo real. Los datos permiten también generar modelos predictivos, indispensables para numerosas empresas. Con *Big Data* veremos el auge de la personalización, por ejemplo en medicina, seguros o *retail*.

Además, *Big Data* no es un compendio de aburridos números o letras, no. El *Big Data* se puede dibujar en forma de bellas visualizaciones. La presentación gráfica de la información se convierte en un elemento esencial a la hora de convertir los datos en negocio.

Y ¿saben qué? Que todo lo que he contado es posible, porque existe 'la nube', ese concepto, a veces difuso, llamado *Cloud Computing*.

En la década de 1980 la revolución informática puso el PC a nuestro alcance. En la década de 1990, la revolución de Internet nos conectó a la Web y a todo el planeta. La entrada en el nuevo siglo nos trajo la revolución de la Web 2.0. Ahora, en esta segunda década del siglo XXI estamos asistiendo a surgimiento de una nueva revolución: la de los datos.

Notas

[1] Véase:
http://www.fundacion.telefonica.com/es/que_hacemos/convocatorias/detalle/06_11_2012_esp_3546

[2] Véase: *http://macrodatos.com/*

[3] Véase: *https://twitter.com/spradas*

Fundación Telefónica
Telefónica

[4] Véase: *https://twitter.com/mtascon*

[5] Véase: *https://twitter.com/sandopen*

[6] Véase: *http://www.farsens.com/*

[7] Véase: *http://www.libelium.com/es/*

[8] Véase: *http://www.panstamp.com/*

[9] Véase: *http://www.cilabweb.com/*

[10] Véase: *http://data.gov.uk/apps*

[11] Véase: *http://www.itu.int/es/Pages/default.aspx*

[12] Véase: *https://cosm.com/*

[13] Véase: *http://www.sorayapaniagua.com/2012/07/23/josep-curto-de-idc-big-data-es-un-cambio-transgresor-y-disruptivo/*

[14] Véase: *https://www.carriots.com/*

[15] Véase: *http://www.cubenube.com/index.php*

Fundación Telefónica
Telefónica