

# **Big data: ¿solución o problema?**

Estrella Pulido Cañabate

Lección Inaugural  
Curso Académico 2014-2015



*Excelentísimo y Magnífico Sr. Rector,*

*Sr. Presidente del Consejo Social,*

*Sra. Secretaria General,*

*Compañeros y compañeras de la comunidad universitaria,*

Cuando el Rector me llamó a últimos de junio para proponerme impartir esta lección inaugural pensé que debía tratarse de una equivocación porque, por lo que yo recordaba, los profesores que la habían impartido en otras ocasiones tenían una larga experiencia de muchos años en la Universidad o eran eminencias en algún tema relevante.

Transcurridos unos minutos empecé a ser consciente del honor que suponía poder hablar delante de una audiencia tan prestigiosa como la que ocupa hoy este salón en el acto de apertura del curso académico de una Universidad a la que me siento muy orgullosa de pertenecer. Por todo esto, quiero dar las gracias al Rector y a su equipo de gobierno por haber pensado en mí y también por sugerirme que hablara de un tema como el Big Data, que me brinda la oportunidad de compartir con ustedes mi visión de que la tecnología tiene mucho que aportar en la construcción de una sociedad en la que todos podamos vivir mejor.

Es la visión que me motivó a estudiar la carrera de Ingeniería Informática hace ya algunos años. En aquella época, no se tenía en muy buen concepto a las nuevas tecnologías porque, según me repetían las personas allegadas, quizá por desconocimiento, sólo servían para hacer desaparecer puestos de trabajo. Yo veía claro que esto podía ser cierto pero sólo para los trabajos más tediosos y repetitivos y que, a cambio, las tecnologías ofrecían numerosas aplicaciones que podían mejorar la sociedad y hacerla avanzar.

Hay quien opina que las mujeres elegimos carreras donde podamos ayudar a otras personas y que, por eso, nos decantamos por la medicina o la docencia y somos una minoría en el mundo de las ingenierías y de la tecnología. Ojalá las mujeres que formamos parte de esa minoría consiguiéramos transmitir a las nuevas generaciones lo que quizá yo entreveía cuando empecé a estudiar: que los problemas que puedan surgir como consecuencia de los avances tecnológicos no se deben a la tecnología en sí, sino al uso que hagamos de ella. Eso es lo que sigue motivándome a trabajar en este campo y espero poder transmitirles con esta lección inaugural.

El Big Data: ¿solución o problema? Mi intención con esta lección no es resolver esta disyuntiva, sino aclarar lo que significa este concepto, describir un mundo que, como refleja el diseño que aparece en la portada de esta lección, está lleno de luces y también de sombras, y presentar una panorámica, tanto de lo que el Big Data puede aportar en la construcción de una sociedad mejor, como de los problemas que puede generar.

Estoy segura de que muchos de ustedes han oído el término “Big Data”, porque es un tema que parece que ahora está en boca de todos. Se habla de él hasta en el fútbol, donde se dice que puede revolucionar la toma de decisiones por parte de los entrenadores. Como no soy muy aficionada a ese deporte, desconocía que se analizan un montón de detalles sobre los partidos, como el número de disparos a puerta que acaban en gol, la distancia desde la que se efectúan o el grado en el que cada jugador depende de sus compañeros para marcar un tanto. Aplicando técnicas de Big Data a todos estos datos, Nate Silver, un estadístico y escritor norteamericano, ha identificado a los dos mejores jugadores de fútbol del mundo, cuyo nombre no revelaré aquí para evitar rivalidades (Pereda y Faus, 2014).

Se habla tanto del Big Data que hay quienes comparan la situación actual de esta tecnología, al menos en España, con el sexo adolescente, porque “todos hablan sobre ello, en realidad nadie sabe cómo

hacerlo, todos creen que todos los demás lo hacen, así que todos aseguran que lo están haciendo” (Leonhard, 2014).

A algunos de ustedes el término Big Data puede simplemente intrigarles. Los que hayan leído o escuchado algo sobre el tema puede que se estén debatiendo entre la esperanza y la preocupación porque les surjan preguntas como ¿ayudará el Big Data a curar enfermedades? ¿O dará lugar a nuevas desigualdades médicas? ¿Contribuirá el Big Data a prevenir suicidios? ¿O se usará para rastrear los movimientos de los manifestantes en las calles?

Vivimos rodeados de teléfonos inteligentes, electrodomésticos inteligentes, sistemas GPS, sensores, medidores, cámaras de circuito de televisión... Todos estos dispositivos electrónicos tienen la capacidad de ejecutar numerosas aplicaciones, de comunicarse entre ellos y de generar cantidades inimaginables de datos. Además, gracias a los avances en las telecomunicaciones, las personas y las cosas estamos casi permanentemente interconectadas y generamos también una gran cantidad de información de forma consciente o inconsciente.

Lo cierto es que la información disponible ha aumentado en los últimos años de forma exponencial.

### **¿De dónde proceden todos estos datos?**

Esta inmensa cantidad de datos procede de fuentes muy diversas. Una parte son datos que se recogen sobre nuestras llamadas telefónicas, transacciones bancarias, pagos con tarjeta o búsquedas en Google. Incluso sobre nuestros movimientos a través de las señales GPS procedentes de nuestros teléfonos móviles. Otros datos los generamos nosotros de forma voluntaria cuando publicamos entradas en blogs, subimos imágenes o vídeos a YouTube, o enviamos mensajes a través de redes sociales como Facebook o Twitter.

La ciencia también genera una gran cantidad de datos en campos como la genómica, la física de partículas, la biología, las ciencias medioambientales, la astronomía o la meteorología.

Otra cantidad importante de datos procede de sensores que monitorizan los objetos, capturan datos sobre ellos y comunican esa información a través de la red. Es lo que se conoce como el **Internet de las Cosas**. Se estima que en 2020 habrá en el mundo 26.000 millones de dispositivos conectados (Gartner, 2014) que generarán el 40% de la totalidad de los datos creados (Capgemini, 2014).

En el mundo de ese año 2020, no es difícil ni inverosímil imaginar una vida como la de Andrés, que en el tren de vuelta a casa está utilizando su teléfono inteligente para consultar su blog favorito de cocina y acaba de decidirse por una receta para la cena de sus dos hijos. Envía una petición a su frigorífico inteligente que, a su vez, confecciona una lista de la compra y la remite al supermercado para asegurarse de que los ingredientes esenciales no estarán agotados cuando Andrés llegue. Como sus hijos no son muy aficionados a ir de compras en el coche cuando salen de la guardería, Andrés envía una petición a su sistema de entretenimiento doméstico para que seleccione un episodio de la serie favorita de televisión de sus hijos y lo descargue en el sistema multimedia del coche. Si fuéramos más allá con el ejemplo, Andrés podría dar una orden al horno para que se precalentara, comprobar si se ha cargado la batería de su coche eléctrico, y quizá también averiguar, con la ayuda de los chips cosidos en la ropa de sus hijos, si tiene que recogerles en el patio o en la sala de música de la guardería (Cramer, 2012).

Esta situación está más cerca de ocurrir de lo que pensamos, porque los circuitos integrados cada vez son más baratos y pueden añadirse sensores a casi todo. Las fábricas, las ciudades y el transporte están transformándose en fábricas inteligentes, ciudades inteligentes y transporte inteligente.

Las **fábricas inteligentes** (Van Rijmenam, 2013a) integran sensores y software en la maquinaria industrial, consiguiendo así optimizar

sus procesos de producción, ahorrar energía y crear mejores productos.

Un ejemplo de **ciudad inteligente** o “smartcity” es Songdo (Lobo, 2014), una ciudad en Corea del Sur que empezó a construirse en el año 2000 y en la que se espera que vivan 65.000 personas en 2016. La infraestructura de esta ciudad contiene sensores que monitorizan y regulan desde la temperatura hasta el consumo de energía y el tráfico. Casi cada dispositivo, edificio o carretera en Songdo está equipado con sensores inalámbricos o microchips, lo que permite, por ejemplo, que el número de farolas encendidas se ajuste automáticamente al número de personas que haya en ese momento en la calle. Además, todas las casas están equipadas con sensores controlados mediante domótica, y el tráfico se regula también mediante sensores instalados en los coches que envían datos de geo-localización y permiten monitorizar las áreas congestionadas de la ciudad.

Un ejemplo que permite ilustrar lo que puede significar el **transporte inteligente** es el de la compañía US Xpress (Van Rijmenam, 2014a) que ha instalado, en cada uno de sus camiones, casi mil sensores que generan datos sobre el consumo de carburante, la velocidad, el estado de los neumáticos o el uso de los frenos. Combinando los datos recibidos de estos sensores con información sobre las condiciones de las carreteras, el estado del tráfico, los datos meteorológicos o la localización de estaciones de servicio, es posible optimizar rutas de envío, reducir el consumo de combustible y la contaminación, aumentar la seguridad en las carreteras y llevar un control de inventario mucho más preciso. Y no sólo los camiones. Los trenes, los aviones, los barcos... Cualquier medio de transporte puede monitorizarse en tiempo real utilizando sensores y navegación por satélite.

En resumen, y simplificando mucho, podemos decir que las principales fuentes de los grandes datos son (1) los datos procedentes de transacciones como compras por Internet o presenciales, transacciones bancarias o búsquedas online; (2) los datos procedentes de

cualquier tipo de máquina, ya sea maquinaria industrial, un equipo de secuenciación de genomas, el acelerador de partículas del CERN, un tomógrafo con el que se realizan TACs o el motor de un avión en vuelo; y (3) los datos compartidos por los usuarios a través de redes sociales (Arcplan, 2012).

### **Pero entonces, esto de almacenar datos ¿es algo nuevo?**

Pues no, almacenar datos no es nuevo. Desde hace mucho tiempo las empresas han venido guardando información acerca de sus clientes, proveedores y operaciones. La administración dispone de enormes bases de datos que contienen información sobre censos de población, registros médicos e impuestos. También se recogen datos cada vez que sacamos dinero en un cajero automático o pagamos con nuestra tarjeta físicamente o a través de Internet.

Pero hasta ahora las empresas no hacían mucho más con esos datos que almacenarlos, porque procesarlos era costoso en tiempo y en esfuerzo.

### **Entonces, ¿por qué ha surgido el Big Data?**

En primer lugar, porque el almacenamiento se ha abaratado enormemente. Hace dos décadas hacía falta una máquina del tamaño de un frigorífico y con un peso de 800 kilos para almacenar un gigabyte de datos. Hoy en día, la mayoría de ustedes lleva, al menos, 16 gigabytes de datos en su teléfono inteligente.

El precio de los dispositivos de almacenamiento también ha bajado. En esas dos décadas de las que hablamos, el coste de almacenar un gigabyte ha pasado de más de 700 euros a cuatro o cinco céntimos.

Además, la velocidad de procesamiento ha aumentado enormemente con la aparición, a principios de los años 2000, de la computación

paralela masiva. En vez de procesar tareas de una en una, ahora los ordenadores pueden procesar una gran cantidad de tareas en paralelo, es decir, todas a la vez. Así es como Google, Facebook o Amazon han sido capaces de construir sus servicios.

No sólo el hardware ha aumentado su velocidad. También ha sido decisiva la aparición de software inteligente que permite sacar partido de la capacidad de procesamiento paralelo, analizando grandes cantidades de datos en muy poco tiempo. Un detalle importante es que este software inteligente puede analizar, no sólo datos almacenados y estáticos, sino también datos volátiles que se analizan en tiempo real a la vez que se producen.

Además de todo esto, han aparecido nuevos tipos de datos que se generan de forma masiva, como correos electrónicos, mensajes de Facebook, tuits o vídeos de YouTube.

La principal diferencia entre estos nuevos tipos de datos y los que se han venido recogiendo hasta ahora es que estos últimos son datos estructurados, es decir, bien organizados, como los que pueden aparecer en una hoja de cálculo o en una base de datos, mientras que los primeros carecen de una estructura identificable. El análisis de estos datos no estructurados es mucho más difícil, pero a la vez también mucho más valioso.

## **¿Y más o menos de cuántos datos estamos hablando?**

Para que sean ustedes capaces de estimar lo grande que es el Big Data, empezaremos por el megabyte, que equivale a 1 millón de bytes y nos permitiría almacenar un libro de unas 500 páginas que tuviera sólo texto (Grant, 2012).

En un terabyte, equivalente a un millón de megabytes, podríamos almacenar 2.767 copias de la Enciclopedia Británica, o 16.667 horas de música, o 1.333 horas de video.

Se necesitarían casi 5 exabytes, o 5 millones de terabytes, para almacenar las secuencias del genoma humano de todas las personas del mundo.

Pues bien, la compañía americana IDC estima que en 2011 se crearon 1,8 zettabytes de información, es decir, 1.800 exabytes, una cantidad de datos con la que podrían llenarse 57.500 millones de iPads de 32 gigabytes. Con este número de iPads podría construirse una pared dos veces más alta que la muralla china.

Por si todavía no se sienten ustedes suficientemente abrumados, en 2013 se generaron 4,4 zettabytes de datos, de los cuales el 49% eran datos no estructurados o semiestructurados. IDC pronostica que en 2020 generaremos 44 zettabytes, es decir, 10 veces más (IDC, 2014; TCS, 2013).

Toda esta cantidad de información se genera a una enorme velocidad, tan alta que el 90% de los datos que existían en el mundo en 2012 se había creado en los dos años anteriores. Cada minuto, por ejemplo, se envían más de 200 millones de mensajes de correo electrónico, se realizan más de 4 millones de búsquedas en Google y se suben 72 horas de vídeos nuevos a YouTube (Gunelius, 2014). En media hora, el motor de un avión que vuela de Londres a Nueva York genera 10 terabytes de datos (MacKinnon, 2013). En 2013 se publicaron al día 400 millones de tuits (Griggs y Kelly, 2013) y se compartieron 4.500 millones de mensajes “me-gusta” en Facebook (Noyes, 2014).

## **Entonces, ¿qué es el Big Data?**

La empresa consultora Gartner define el Big Data como “aquellos recursos de información caracterizados por su alto volumen, velocidad o variedad, que requieren formas de procesamiento innovadoras y eficientes para la mejora del conocimiento y la toma de decisiones”<sup>1</sup>.

---

<sup>1</sup> <http://www.gartner.com/it-glossary/big-data/>

De acuerdo con esta definición, las características que definen el Big Data pueden resumirse en lo que se conoce como las tres Vs: **Volumen**, **Velocidad** y **Variedad** (Laney, 2011).

A estas tres Vs que forman parte de la definición intrínseca del Big Data, los expertos añaden una cuarta V que tiene que ver con la **Veracidad**, ya que es fundamental que los datos en los que se basan los análisis sean correctos y no contengan sesgos o ruido que puedan distorsionar las conclusiones que se extraigan de ellos (Normandeau, 2013).

La definición de Big Data propuesta por Gartner tiene una segunda parte, también importante, que tiene que ver con la capacidad para analizar los datos y extraer de ellos información relevante. Andreas Weigend, antiguo científico de Amazon y profesor en varias universidades americanas, afirma que los datos son el nuevo petróleo, no sólo en el sentido económico, sino también porque, al igual que el petróleo, es necesario refinarlos y depurarlos para que aporten valor (Brustein, 2014a).

Adoptando la definición propuesta por Gartner, en el resto de esta lección usaremos el término Big Data para referirnos de manera indisoluble a los grandes conjuntos de datos y a los resultados que puedan derivarse del análisis de los mismos.

## **¿Qué cambios está provocando el Big Data?**

El Big Data está causando una revolución en el mundo empresarial, modificando de manera sustancial los negocios existentes y creando otros completamente nuevos.

IDC estimó en 2012 que la tecnología y el mercado de servicios del Big Data crecerán a un ritmo anual del 31,7% hasta 2016. Esto es siete veces más que el ratio de crecimiento conjunto de todas las Tecnologías de la Información y la Comunicación (TIC) (IDC, 2012).

Como consecuencia del Big Data, se estima que para el 2015 se habrán creado en el sector de las TIC 4,4 millones de puestos de trabajo en todo el mundo, de los cuales 1,2 millones se crearán en Europa (Gartner, 2012).

El problema es que no habrá suficientes especialistas para cubrir esta demanda. Se estima que en el 2018, sólo en Estados Unidos puede haber una carencia de entre 140.000 y 190.000 profesionales con conocimientos profundos sobre técnicas de análisis (McKinsey, 2011) y capaces de presentar, muchas veces de forma visual, los resultados de los análisis a los ejecutivos responsables de tomar las decisiones. El desarrollo de técnicas de visualización específicamente diseñadas para los grandes conjuntos de datos será fundamental, porque permitirán presentar la información de una forma gráfica e intuitiva (Davenport y Dyché, 2013).

Según un informe presentado en 2013 por Interxion, una empresa que gestiona más de 30 centros de datos en 11 países de Europa, basado en entrevistas a 750 ejecutivos europeos del área de las TIC, en el caso de España el 83% de las empresas está convencido de que el Big Data se convertirá en una prioridad en los próximos tres años (Interxion, 2013).

Teniendo en cuenta todos estos datos, está claro que la figura del científico de datos será indispensable para las empresas, por lo que es fundamental que las Universidades se planteen como una prioridad la formación de este tipo de perfil profesional.

### **¿Para qué se usa el Big data?**

Los más optimistas opinan que en la que parece inminente era del Big Data lograremos una mejora sustancial de nuestra capacidad para realizar diagnósticos y pronósticos fiables en numerosas áreas de la vida social.

Esta nueva capacidad de análisis y predicción puede ser aprovechada por las empresas con fines puramente lucrativos utilizando distintas técnicas, como la publicidad personalizada o el análisis de sentimientos.

## **La publicidad personalizada**

Localizar patrones y tendencias permite a las empresas de comercio online adecuar los productos y servicios a los clientes, anticipar la demanda o mejorar las ventas a través de incentivos, como descuentos, envíos gratuitos o facilidades de pago.

Estas empresas analizan los datos que se almacenan cuando navegamos por Internet, viajamos con nuestros móviles, hacemos compras con tarjetas de crédito o comunicamos nuestras actividades, objetivos y preferencias a través de las redes sociales, con el fin de recomendarnos los productos que mejor se adaptan a nuestras necesidades y conseguir prolongar al máximo nuestra interacción con la página web de su comercio (McKinsey, 2013). Un buen ejemplo es Amazon, que analiza los datos de sus 152 millones de clientes para utilizarlos en su motor de recomendación.

Pero no sólo el comercio online puede obtener beneficios del Big Data. En centros comerciales y aeropuertos se está empezando a usar la **publicidad basada en la localización** (Arora, 2013), que consiste en analizar la información recogida a través de la red wifi del centro comercial para localizar las tiendas en las que un cliente pasa más tiempo, o los productos por los que muestra más interés. Los comercios combinan estos datos con otros relacionados con el poder adquisitivo o el historial de compras de los clientes para personalizar la publicidad y enviar a sus teléfonos móviles anuncios y promociones adaptadas a sus necesidades.

El Big Data y la analítica predictiva van más allá de las recomendaciones y la publicidad personalizadas. Google Now ofrece, desde

2013, una nueva funcionalidad de lo que se conoce como **sistemas anticipativos** (Regalado, 2013). Estos sistemas tienen una filosofía similar a los sistemas de recomendación, pero no esperan a que expresemos lo que queremos, en forma de consulta o petición, sino que se adelantan y nos hacen sugerencias teniendo en cuenta la hora del día, el tiempo que hace, nuestras conexiones sociales, las reuniones anotadas en nuestra agenda o el estado del tráfico.

Las sugerencias pueden incluir un aviso sobre novedades en alguna de nuestras webs o blogs favoritos, una notificación cuando pasamos por una tienda que dispone de alguno de los productos que hemos buscado recientemente en Internet, una estimación del tiempo que tardaremos en llegar a una cita teniendo en cuenta el estado del tráfico, o el tiempo que hace en el destino de nuestro próximo viaje.

### **Análisis de sentimientos**

Analizar las enormes cantidades de datos que se mueven en las redes sociales también puede ser de gran utilidad para las empresas, sobre todo teniendo en cuenta el número de usuarios cada vez mayor que interacciona en ellas. Facebook, por ejemplo, tiene más de mil millones de usuarios, Twitter 500 millones, y LinkedIn 225 (Brannock, 2014).

Gracias a estos análisis, las empresas pueden entender lo que el público opina sobre distintos productos, servicios, anuncios publicitarios o series de televisión, y saber quiénes son los usuarios con más influencia en las redes. De esta forma pueden mejorar sus productos o incluso desarrollar otros nuevos, de acuerdo con el deseo de los consumidores. Es lo que se conoce como el análisis de sentimientos o minería de opiniones (Van Rijmenam, 2013b).

En la industria del entretenimiento, el resultado de estos análisis puede combinarse con otros datos, como el número de veces que los

espectadores paran, avanzan o retroceden al ver una serie de televisión online, para predecir si una película o una serie tendrán éxito.

## **El Big Data y el gobierno**

Como hemos visto, los datos recogidos sobre los clientes pueden ser utilizados por las empresas para mejorar la experiencia del consumidor y, en último término, para incrementar las ganancias. Los Gobiernos también pueden usar estas herramientas para el espionaje masivo (Naughton, 2014) o para predecir resultados electorales como el sitio web FiveThirtyEight<sup>2</sup> que en 2008 logró predecir correctamente el vencedor de 49 de los 50 Estados norteamericanos en las elecciones presidenciales que ganó Barack Obama.

Pueden utilizarlo también para reducir costes, por ejemplo, en el sector sanitario. Como veremos más adelante, mediante la predicción y la personalización pueden ahorrarse costes en hospitalización y en tratamientos. El estudio de 2011 del McKinsey Global Institute calcula que la aplicación del Big Data en el sector sanitario podría representar un ahorro de hasta 300.000 millones de dólares al año para el sistema de salud pública en Estados Unidos (McKinsey, 2011).

Pero, ¿qué pasaría si los Gobiernos utilizaran los datos sobre nuestras búsquedas en Google o nuestras interacciones en las redes sociales para proporcionar un mejor servicio a los ciudadanos y abordar problemas relacionados con la atención sanitaria, el desempleo, los desastres naturales o el terrorismo?

Veamos algunas de las iniciativas que los Gobiernos están llevando a cabo en este sentido. En agosto de 2012, el Servicio Geológico de Estados Unidos rastreó miles de tuits buscando la palabra “terremoto”. Utilizando datos sobre la hora y la posición contenidos en

---

<sup>2</sup> <http://fivethirtyeight.com/>

esos tuits consiguió localizar un gran terremoto en las islas Filipinas antes de que los sismógrafos lo registraran (Konkel, 2013a).

Un método parecido puede utilizarse para predecir variaciones en la tasa de desempleo. Por ejemplo, Global Pulse, una iniciativa de Naciones Unidas para abordar crisis socioeconómicas, en colaboración con la empresa SAS, analizó conversaciones que tenían lugar en redes sociales de Estados Unidos e Irlanda y consiguió predecir aumentos en la tasa de desempleo tres meses antes que los informes oficiales (Cary, 2012).

Con el Big Data se pueden reducir también los índices de criminalidad. Por ejemplo, en la ciudad de Los Ángeles se usa para detectar en qué calles, o qué grupos o individuos son más propensos a cometer delitos y deben ser sometidos a una vigilancia extra (Van Rijmenam, 2014b).

Puede ayudar a resolver delitos como las dos explosiones que mataron a tres personas e hirieron a otras muchas en el maratón de Boston de 2013. Menos de 24 horas después de que ocurrieran, el FBI había recogido 10 terabytes de datos que incluían registros de llamadas, mensajes de texto, fotografías y videos procedentes de cámaras de vigilancia. La investigación concluyó cuatro días después con la muerte de uno de los sospechosos y la detención del segundo, en parte gracias al análisis de todos estos datos y de la colaboración ciudadana, que contribuyó con pistas, fotografías y videos (Konkel, 2013b).

El Departamento de Seguridad de Estados Unidos defiende que el Big Data puede utilizarse para prevenir el terrorismo y está trabajando en el proyecto FAST (Future Attribute Screening Technology), que tiene como objetivo identificar a posibles terroristas monitorizando constantes vitales, lenguaje corporal y otros patrones fisiológicos (Armunanto, 2013). Este uso puede ser cuestionable y controvertido como refleja la interesante novela “Little brother” de Cory Doctorow que trata de los peligros asociados al mal uso de estas técnicas (Doctorow, 2007).

El Big Data también puede ser de gran ayuda en el seguimiento de epidemias como demostró Google en 2009, descubriendo que existía una estrecha relación entre el número de personas que realizaba búsquedas relacionadas con la gripe y las personas que realmente sufrían síntomas gripales. En la actualidad, esta herramienta, conocida como Google Flu Trends, ofrece datos sobre la actividad de la gripe en diferentes países y regiones de todo el mundo.

En el hospital de niños de Boston, en Estados Unidos, los investigadores han desarrollado un método parecido, pero a partir del número de visitas a páginas de la Wikipedia. Este modelo estima los niveles de gripe hasta dos semanas antes de que estén disponibles los datos oficiales, y estima la semana de mayor actividad de la gripe un 17% mejor que Google Flu Trends (Brustein, 2014b).

El análisis de datos también puede utilizarse en el sector agrario para mejorar la producción agrícola, minimizar su impacto medioambiental y reducir costes mediante lo que se conoce como **agricultura de precisión**, que combina datos meteorológicos con los de temperatura y humedad del terreno o el grado de luminosidad recogidos mediante sensores para ayudar a los agricultores a tomar mejores decisiones con respecto a la siembra, fertilización y cosecha de sus cultivos (Mello y Treinish, 2012).

Como vemos, la mayoría de estas iniciativas proceden de Estados Unidos, donde la administración de Obama está dedicando 200 millones de dólares a la Iniciativa para la investigación y el desarrollo en Big Data, un programa anunciado en marzo de 2012 cuyo objetivo es explorar cómo pueden utilizarse estas herramientas para afrontar los problemas más urgentes del gobierno (Kalill, 2012).

En Europa, de momento, el Big Data sólo puede vislumbrarse como una oportunidad de futuro, aunque en julio de este año la Comisión Europea hizo una llamada a los gobiernos para que “despierten a la revolución del Big Data” (Kroes, 2014) y propuso la creación de diversas iniciativas dentro del programa Horizonte 2020, un Programa

Marco de 70 billones de euros para actividades de investigación e innovación en el período 2014-2020.

## **El Big Data y el desarrollo**

La revolución de los datos no se limita al mundo industrializado y está ocurriendo también en los países en vías de desarrollo. En países de África y del sur de Asia, el 50% de los adultos son propietarios de un teléfono móvil, y entre un 10 y un 20% más, aunque no son propietarios, tienen acceso al móvil de amigos, familiares o vecinos (Cartesian, 2014).

Quizá muchos de ustedes ya sepan que cada vez que hacemos una llamada con nuestro teléfono móvil se genera un registro digital o CDR que incluye información como la hora de la llamada, la duración y la identificación de la antena a la que se conectaron los teléfonos del emisor y el receptor en el momento de la llamada. Como el operador de la red conoce las posiciones de estas antenas, a partir de esta información, es posible conocer la ubicación de las dos partes. Esta información geoespacial es extremadamente útil para programas de ayuda humanitaria y al desarrollo (Global Pulse, 2013).

Analizando estos datos se pueden reconstruir, por ejemplo, los patrones de movimiento de una comunidad de personas y modelizar, entre otras cosas, la propagación de enfermedades. Es lo que se hizo en Kenia para localizar lugares a donde la enfermedad de la malaria tenía una probabilidad más alta de propagarse.

El análisis de CDRs se utilizó también para organizar la ayuda humanitaria a los afectados por el terremoto de Haití de 2010, o para medir la efectividad de las restricciones de movilidad que impuso el gobierno mejicano a los ciudadanos para controlar la propagación del virus de la gripe H1N1 en 2009.

Un último ejemplo de cómo se puede utilizar el análisis de la información geoespacial para gestionar situaciones de crisis es la plataforma Ushahidi que, utilizando informes enviados por los ciudadanos a través de la web y de teléfonos móviles, creó un mapa interactivo donde se reflejaban los incidentes violentos que ocurrieron en Kenia después de las elecciones de 2008 (Jeffery, 2011).

Como vemos, en la ayuda al desarrollo, al igual que en otras áreas, el análisis de los datos también supone una mejora en la toma de decisiones. Esto adquiere un significado aún mayor en países en vías de desarrollo donde el impacto de estas decisiones afecta a las vidas de las personas, y las medidas que se tomen como resultado de las mismas pueden resultar cruciales.

## **El Big Data y la salud**

Un campo importante de aplicación del Big Data es el de la salud y la atención sanitaria. Hoy en día se almacenan todo tipo de datos sobre pacientes, enfermedades, tratamientos, medicación y resultados. Nuestro cuerpo se ha convertido en una fuente más de datos. Radiografías, mamografías, TACs, resonancias magnéticas... Se calcula que, para 2015, un hospital medio en Estados Unidos gestionará 665 terabytes de datos frente a los 168 que gestionaba en 2010 (Iron Mountain, 2013). El análisis de todos estos datos juega un papel fundamental en lo que, según los expertos, será la medicina del futuro o medicina “4P” (Hood y Galas, 2008): una medicina personalizada, predictiva, preventiva y participativa. Veamos cómo el Big Data puede aplicarse en cada una de estas áreas.

### **Personalización**

El análisis de datos genéticos puede desempeñar un papel relevante en el campo de la medicina personalizada, ya que permite realizar diagnósticos más precisos y diseñar terapias más enfocadas a cada paciente.

Gracias a la secuenciación paralela o de siguiente generación, conseguir estos datos es cada vez menos costoso y más rápido. Mientras que se necesitaron 13 años para conseguir la secuencia completa de un genoma humano con un coste de unos 2.047 millones de euros, ahora hay equipos que permiten secuenciar el ADN de una persona en solo un día y por menos de 800 euros (Prats, 2012).

El análisis de datos genéticos tiene dos áreas de aplicación fundamentales. En primer lugar, las técnicas de Big Data pueden ayudar a entender cómo funcionan los genes y cómo interactúan entre ellos para prevenir o luchar contra enfermedades como el cáncer.

En segundo lugar, analizar los datos genéticos de cada individuo permite personalizar los tratamientos médicos y entender mejor el desarrollo de una enfermedad en un individuo concreto. Hay empresas que ya están empezando a predecir qué medicamentos contra el cáncer son más efectivos para un determinado perfil genético y tipo de cáncer, lo que aumenta su eficacia y reduce los efectos secundarios no deseados.

Gracias al Big Data, también pueden estudiarse los efectos secundarios perjudiciales de la medicación, así como las interacciones entre distintos medicamentos, y utilizar esta información para configurar tratamientos a medida que optimicen los resultados y minimicen los riesgos.

## **Predicción**

Para ilustrar cómo el Big Data puede influir en la medicina predictiva me gustaría que nos situáramos por un momento en la sala de cuidados intensivos de neonatos del hospital infantil de Toronto, donde sólo se escuchan los pitidos de los monitores que registran, entre otras cosas, los latidos del corazón de los bebés. Las enfermeras a cargo de esta sala saben que pequeños cambios en el ritmo cardiaco o en la respiración de los bebés pueden indicar un empeoramiento,

pero a veces estos cambios pueden ser tan sutiles como para no disparar ninguna alarma en el monitor y pasar por ello desapercibidos.

Pero gracias al Big Data estos signos de alarma pueden detectarse en cuanto aparecen. Esto es lo que hizo una profesora de informática del Instituto de Tecnología de la Universidad de Ontario. Con la ayuda de Watson, el supercomputador de IBM que derrotó a los campeones del concurso de televisión americano Jeopardy! en 2011, analizó en tiempo real los datos sobre el ritmo cardiaco de los bebés recogidos por el monitor y encontró que los patrones detectados revelaban signos de infección 24 horas antes de que los bebés mostraran algún síntoma visible. En bebés prematuros, adelantar un tratamiento aunque sólo sea en una hora puede ser cuestión de vida o muerte (Dembosky, 2012).

## **Prevención**

Un ejemplo de cómo se puede utilizar el Big Data para conseguir una medicina preventiva es el trabajo que están llevando a cabo investigadores de la Escuela de Enfermería Sinclair en la Universidad de Missouri. Gracias al análisis de los datos recogidos por sensores de movimiento instalados en las camas, paredes y suelos de las viviendas de gente anciana, estos investigadores están identificando patrones de agitación durante el sueño y de cambios en la actividad física diaria de estas personas mayores que pueden prevenir posibles caídas y visitas a urgencias (Cimons, 2012).

También es posible prevenir ataques al corazón o comas diabéticos analizando los datos generados por dispositivos de monitorización de uso doméstico, como medidores de glucosa o monitores cardiacos.

## **Participación**

La medicina será participativa cuando las personas tomemos conciencia de que somos los máximos responsables de nuestra propia

salud y nuestro mejor médico. En ese sentido, el Big Data tiene mucho que aportar, porque permite pasar de los cuidados preventivos a los cuidados proactivos.

Cada vez hay más dispositivos “ponibles” y aplicaciones móviles que hacen seguimiento de nuestros hábitos de salud (Wall, 2014). Estos dispositivos forman parte del movimiento que se conoce como el “yo cuantificado”, que consiste en monitorizar el cuerpo y el comportamiento con el objetivo de mejorar la salud, el estado físico o la productividad.

## **El Big Data y la educación**

Dicen que lo que no se puede medir no se puede mejorar, y si hay un área que puede beneficiarse del análisis de los grandes conjuntos de datos es el de la educación. Al igual que las empresas y la Administración, la Universidad viene recogiendo datos sobre los estudiantes desde hace mucho tiempo, la mayoría de ellos relacionados con su rendimiento escolar. Se almacenan las calificaciones de los trabajos que entregan, sus resultados en los exámenes, el número de convocatorias que necesitan para superar una asignatura o cuánto tiempo tardan en finalizar sus estudios.

La novedad es que las universidades cada vez ofrecen más materiales educativos a través de plataformas de e-learning como Moodle, y el aprendizaje online a través de los MOOCs o Massive Open Online Courses cada vez está más extendido. Estos cursos online ofrecen numerosas ventajas, porque permiten a los estudiantes acceder a la enseñanza en el momento y lugar preferidos y les proporcionan un feedback inmediato y constante sobre su rendimiento, que influye muy positivamente en la motivación.

Plataformas de enseñanza online, como edX, Coursera o Udacity, hacen posible que 100.000 estudiantes repartidos por todo el planeta

puedan asistir a una clase de un profesor de la Universidad de Harvard. Pero el gran impacto que pueden tener los MOOCs en la educación no tiene que ver sólo con esto, sino también con los conjuntos enormes de datos sobre el comportamiento de los estudiantes que recogen estas herramientas, cuyo análisis puede ayudar a mejorar el rendimiento académico de los estudiantes, a disminuir las tasas de abandono y a personalizar la educación adaptando los contenidos, las tareas y el feedback a las necesidades de cada estudiante (Guthrie, 2013; Nielson, 2013).

Analizar toda la variedad de datos que se recogen en el proceso de formación de los estudiantes es lo que se conoce como Analítica del Aprendizaje (Ferguson, 2014). Estos datos pueden tener que ver con el rendimiento académico, pero también con la interacción de los estudiantes con el campus universitario en general, como el número de accesos a la plataforma docente virtual, el uso de la biblioteca, la participación en actividades deportivas, la asistencia a reuniones con el tutor, o incluso el uso que hacen del aparcamiento.

Este análisis puede servir a los estudiantes individuales para reflexionar sobre sus resultados y modelos de comportamiento en relación con otros compañeros; a los profesores para identificar qué estudiantes tienen más riesgo de abandonar o fracasar y necesitan más apoyo y atención; a los encargados de la calidad docente para introducir mejoras en las asignaturas o desarrollar nuevos planes de estudio; y a los administradores para tomar decisiones sobre temas relacionados con la promoción de los estudios, la distribución de los recursos o el proceso de admisión.

Como todo, el uso del Big Data en el entorno universitario también puede ser controvertido (Renton, 2014). Una de las principales críticas que se plantea con respecto a la analítica del aprendizaje es que puede predecir los resultados académicos que obtendrá un estudiante, pero no sirve para identificar las causas de su éxito o fracaso.

Otro conjunto de críticas tiene que ver con el hecho de que el análisis de datos identifique perfiles de estudiantes que reúnen determinadas características. La primera de ellas es que el conocimiento de estos perfiles por parte de los profesores puede sesgar sus expectativas sobre los estudiantes. Es lo que se conoce como el efecto Pigmalión, o la teoría de la profecía auto cumplida, que puede definirse como una predicción que, una vez hecha, es en sí misma la causa de que se haga realidad (Díaz, 2014).

Por otro lado, si estos perfiles se utilizan pensando sólo en el beneficio económico, identificar el perfil de los estudiantes más propensos al abandono puede provocar que no se les admita, o que se les asignen menos recursos porque resultan menos rentables.

## **El Big Data y la investigación**

La computación intensiva de datos juega un papel fundamental en el descubrimiento científico. Hallazgos como el Bosón de Higgs, aunque pendiente de la confirmación definitiva, o la secuenciación del genoma humano no habrían sido posibles sin el Big Data.

Los científicos están acostumbrados a los grandes conjuntos de datos que puede generar, por ejemplo, un secuenciador de ADN, un acelerador de partículas, un telescopio o las estaciones y satélites meteorológicos.

El Big Data permite entender mejor el flujo de tráfico en las autopistas, el uso de energía doméstica, las condiciones en el interior de los volcanes, los huracanes, las enfermedades o el universo (CRA, 2011).

También se puede utilizar para estudiar las variedades dialectales del español como han hecho dos investigadores del Instituto de Física Interdisciplinar y Sistemas Complejos a partir de 50 millones de tuits geo localizados escritos en español recogidos durante dos años (Gonçalves y Sánchez, 2014).

En economía, el Big Data puede utilizarse, por ejemplo, para estimar el índice de inflación. Es lo que hace el proyecto BPP (Billion Prices Project) del MIT a partir de los precios de más de 50.000 productos recogidos diariamente de cientos de tiendas online (Einav y Levin, 2013).

Otra área en la que el Big Data está teniendo un gran impacto es la ciencia social computacional, una disciplina a medio camino entre la psicología y la sociología de la que ya se hablaba en 2009 (Lazer et al., 2009) y que analiza los datos de nuestras interacciones en Internet para revelar patrones de comportamiento individual y grupal.

Ninguno de ustedes se sorprendería si les dijera que el análisis de grandes cantidades de datos se utiliza en disciplinas como la física de partículas, las ciencias medioambientales, la bioinformática o la microbiología. Pero quizá les suenen menos familiares disciplinas como la cliometría, que aplica la econometría a la historia; la estilometría, que es el estudio del estilo de escritura de un autor; o la culturómica, que describe las investigaciones cuantitativas en ciencias sociales y humanidades. En todas estas disciplinas el Big Data tiene mucho que decir.

Como vemos, el Big Data no sólo tiene aplicación en la industria o en la investigación científica, sino también en campos más relacionados con las humanidades, como la sociología, las ciencias políticas o la economía. Las referencias a las aplicaciones del Big Data en todas estas disciplinas son tan numerosas que sería imposible abarcarlas en esta lección, pero sí me van a permitir que mencione unos cuantos detalles generales, aplicables a todas ellas.

En primer lugar, el cambio que puede suponer para la investigación la capacidad del Big Data para encontrar relaciones entre datos que, a primera vista, parecen no tener nada en común. Esto puede ofrecer nuevas perspectivas para la investigación porque, aunque hasta ahora los datos han servido para probar hipótesis, en el futuro pueden ser los datos los que generen las hipótesis.

Otro aspecto importante es la conveniencia de implementar formatos que permitan compartir datos entre distintas disciplinas, aunque a veces sea difícil combinar datos de distintas fuentes, con estructuras a veces incompatibles.

Un último aspecto que me gustaría comentar con ustedes es el del crowdsourcing, un término acuñado en 2006 que puede definirse como “un tipo de actividad participativa online a través de la cual una persona, una institución, una organización sin ánimo de lucro o una empresa proponen a un grupo de individuos, mediante una convocatoria abierta y flexible, la realización libre y voluntaria de una tarea” (Pardo, 2014). Cuando la tarea a realizar tiene que ver con la captura, sistematización o análisis de grandes cantidades de datos, hablamos de ciencia ciudadana y de científicos ciudadanos voluntarios que recogen o procesan datos como parte de una investigación científica (OpenScientist, 2011).

Algunos científicos están usando esta técnica en el campo de la investigación médica, donde los experimentos controlados pueden ser caros e inefectivos, o en áreas donde los datos a analizar tienen formatos que los ordenadores no son todavía capaces de procesar (Ridley, 2012).

Existen numerosas iniciativas relacionadas con la ciencia ciudadana. Una de ellas es Foldit, un juego de ordenador desarrollado en la Universidad de Washington mediante el cual aficionados a este tipo de juegos aportan soluciones a problemas de plegamiento de proteínas. Los astrónomos también involucran a aficionados en la búsqueda de galaxias. Lo mismo ocurre con empresas como 23andme<sup>3</sup>, que se dedica a los análisis genéticos y compara el ADN de los clientes que aportan voluntariamente sus resultados, con el objetivo de encontrar mutaciones que aumentan o disminuyen el riesgo de contraer Parkinson.

La ciencia de los ciudadanos puede ser difícil de adoptar por nosotros los académicos, que estamos acostumbrados a guardar nuestros

---

<sup>3</sup> <https://www.23andme.com/>

datos bajo llave, porque nuestra reputación depende de que seamos los primeros en publicar resultados novedosos.

Quizá también tengamos reticencias con respecto a la fiabilidad de los resultados cuando el análisis se deja en manos de la “multitud”. Pero aunque a los científicos pueda no gustarnos compartir, sí que estamos acostumbrados a colaborar, al menos con otros científicos. ¿Por qué no hacerlo con los ciudadanos? Podría ser una forma de acercar la ciencia a la gente, y quizá supondría una mejora de la interacción entre la ciencia, la sociedad y las políticas de investigación.

## **El Big Data y el gobierno abierto**

No podemos hablar de los grandes datos sin hablar también de los datos abiertos, que la red para el Conocimiento Abierto define como “cualquier contenido, información o datos que las personas pueden usar, reusar y redistribuir sin ninguna restricción legal, tecnológica o social”<sup>4</sup>.

Todas las definiciones de datos abiertos incluyen, como características básicas, que deben estar disponibles de forma gratuita o a un coste mínimo en un formato que facilite su uso, y deben poder ser utilizados sin discriminación de personas, grupos o aplicaciones.

Podemos citar muchos ejemplos de datos que son o deberían ser abiertos. Por ejemplo, los datos sobre títulos y autores de obras culturales que se guardan en galerías de arte, bibliotecas, archivos y museos; los relacionados con la investigación científica, con los presupuestos del Estado o con los mercados financieros; los datos estadísticos como el censo o los indicadores socioeconómicos; datos meteorológicos, climatológicos o medioambientales relativos a la presencia y el nivel de contaminantes, o sobre la calidad de ríos y playas...

---

<sup>4</sup> <https://okfn.org/opendata/>

Muchos de estos datos, como los procedentes de la investigación científica, las bases de datos meteorológicas o el censo, pueden ser clasificados dentro del conjunto de los grandes datos y, al igual que éstos, su análisis puede beneficiar a las empresas, las organizaciones y el público en general, porque les permite crear nuevos negocios, detectar patrones y tendencias y tomar decisiones informadas (Gurin, 2014). Por el contrario, los datos que no están abiertos al público sólo benefician a aquellos que los controlan.

Por ello, el gobierno, que tiene la capacidad y los fondos para recoger grandes cantidades de datos, tiene también la responsabilidad de transformar en datos abiertos aquellos subconjuntos que puedan ser más beneficiosos para los ciudadanos y que les permitan participar en las decisiones que les afectan de forma continua y no sólo depositando su voto en una urna cada cuatro años.

Así queda reflejado en la Ley de Transparencia, Acceso a la Información Pública y Buen Gobierno aprobada en diciembre de 2013, cuyo preámbulo dice textualmente que “La transparencia, el acceso a la información pública y las normas de buen gobierno deben ser los ejes fundamentales de toda acción política. Sólo cuando la acción de los responsables públicos se somete a escrutinio, cuando los ciudadanos pueden conocer cómo se toman las decisiones que les afectan, cómo se manejan los fondos públicos o bajo qué criterios actúan nuestras instituciones podremos hablar del inicio de un proceso en el que los poderes públicos comienzan a responder a una sociedad que es crítica, exigente y que demanda participación” (BOE, 2013).

Este párrafo describe perfectamente el concepto de gobierno abierto, cuyos pilares fundamentales son la transparencia, la participación y la colaboración, (Preciado, 2012) y que Javier Llinares, uno de los autores más referenciados en este tema, define como “aquel que entabla una constante conversación con los ciudadanos con el fin de oír lo que ellos dicen y solicitan, que toma decisiones basadas en sus necesidades y preferencias, que facilita la colaboración de

los ciudadanos y funcionarios en el desarrollo de los servicios que presta y que comunica todo lo que decide y hace de forma abierta y transparente” (Llinares, 2007).

Hasta ahora hemos hablado de las soluciones, las luces y las esperanzas en el Big Data. Nos toca ahora plantear la otra cara de la moneda y hablar de sus limitaciones, de los daños que puede causar al medioambiente, y de cuestiones éticas relacionadas con la privacidad, la transparencia, la pérdida de identidad, la discriminación o la exclusión.

### **¿Para qué no sirve el Big Data?**

Como decíamos al comienzo, lo importante no son los datos, sino las conclusiones que se obtienen al analizarlos. No hay duda de que el Big Data es una herramienta muy útil para detectar correlaciones, especialmente correlaciones sutiles que pueden perderse al analizar conjuntos de datos más pequeños, pero ¿qué limitaciones tiene? ¿Para qué no sirve el Big Data?

La primera de ellas es que el Big Data no permite detectar comportamientos individuales o minoritarios que se salgan de la media, sino tendencias y patrones. Dicho de otra forma, el Big Data no es capaz de encontrar una aguja en un pajar, aunque sí podría decirnos qué longitud y tonalidad tienen la mayoría de los trozos de paja.

Una segunda limitación del Big Data es que puede detectar correlación, pero no causalidad. El riesgo de correlaciones espurias, es decir, de asociaciones estadísticamente robustas que ocurren por casualidad, aumenta cuanto mayor sea la cantidad de datos que se analice (KNC, 2014). Por ejemplo, un análisis de datos recogidos entre 2006 y 2011 puede revelar una correlación entre el número de asesinatos ocurridos en Estados Unidos durante ese período y la cuota de mercado del navegador Internet Explorer. Pero, aunque ambas cifras bajaron en ese período, es difícil imaginar que pueda existir una relación causal entre ellas (Marcus y Davis, 2014).

Ya hemos mencionado que el Big Data puede sugerir hipótesis e indicar a los investigadores dónde iniciar un análisis, pero para crear modelos siempre será necesario un conocimiento profundo de la materia que sólo poseen los expertos. Ellos y los científicos de datos, de los que ya hemos hablado, son imprescindibles para orientar los análisis e interpretar sus resultados.

Un problema del Big Data es que los grandes conjuntos de datos contienen una gran cantidad de ruido, es decir, información inútil e irrelevante, y a veces hasta engañosa, que está creciendo a un ritmo más rápido que los datos realmente válidos (Poynter, 2014).

Otra limitación del Big Data es que los resultados de los análisis tienen un período de validez determinado, debido a que los datos en los que se basan siguen cambiando y es necesario reanalizarlos para tener en cuenta dichos cambios. Un buen ejemplo son las predicciones meteorológicas, cuya calidad ha mejorado enormemente en los últimos 25 años, gracias sobre todo a la mejora de las técnicas de análisis, pero que siguen sin poder predecir de forma fiable con más de una semana de adelanto.

Por último, es importante destacar que no todo se puede predecir. Como señala Nate Silver, existen sistemas caóticos, fenómenos que no se pueden medir o técnicas como la extrapolación que quedan fuera del alcance del Big Data (Poynter, 2014).

## **El Big Data y el medioambiente**

Como comentábamos al principio de esta lección, se prevé que en 2020 generaremos 44 zettabytes de datos. Las empresas que almacenan y procesan toda esta información utilizan grandes centros de datos que consumen enormes cantidades de energía eléctrica. Según un estudio de Jonathan Koomey (Koomey, 2011), en el año 2010 estos centros de datos consumieron entre el 1,1% y el 1,5% de la energía eléctrica mundial. Este consumo está creciendo, aunque a un

ratio cada vez más lento, y el aumento ha pasado de un 100% entre 2000 y 2005 a un 56% entre 2005 y 2010.

Además de que el crecimiento en la demanda de energía cada vez es más lento, la eficiencia energética de los dispositivos cada vez es mayor, lo que quiere decir que consumen menos (Kooimey, 2012).

Esta mejora en la eficiencia ha permitido la creación de portátiles, teléfonos inteligentes, sensores inalámbricos y otros dispositivos móviles. Mientras que un ordenador de sobremesa consume entre 150 y 200 kilovatios hora de electricidad al año, un ordenador portátil necesita entre 50 y 70 kilovatios hora; una tableta, 12 kilovatios hora; y un teléfono inteligente, sólo 2 (Clancy, 2013).

En general, la eficiencia eléctrica de la computación, es decir, el número de cálculos que pueden realizarse por kilovatio hora de energía consumida, se ha duplicado cada año y medio desde el nacimiento de la informática. Esto quiere decir, por ejemplo, que si un ordenador portátil como el MacBook Air funcionara con la eficiencia energética de los ordenadores de 1991, su batería duraría apenas unos 2,5 segundos (Kooimey, 2012).

Otro debate con respecto al consumo energético del Big Data es el que tiene que ver con la computación en la nube, que ofrece a las empresas la posibilidad de usar centros de datos centralizados fuera de sus instalaciones. Teniendo en cuenta que algunas empresas todavía desperdician entre el 10 y el 20% de la capacidad de sus servidores (Clancy, 2013), ¿puede decirse que la computación en la nube es menos contaminante que los servidores locales tradicionales?

Algunos expertos creen que si este modelo fuese adoptado por una mayoría de empresas podría disminuir el consumo energético, aunque otros opinan que si aumenta el uso de la computación en la nube porque el coste es más bajo, esto llevará consigo un mayor uso de las redes y, como consecuencia, un aumento del consumo de energía (Mills, 2013).

Además de la eficiencia energética, las empresas deben preocuparse también de las fuentes de electricidad que eligen (Greenpeace, 2011). Muchas empresas tecnológicas están optando por utilizar en sus fábricas fuentes de electricidad sucias, como el carbón o la energía nuclear. Esto puede tener un gran impacto negativo en nuestro entorno, como indica la estimación de que el 17% de la huella de carbono de la tecnología a nivel mundial se debe a los centros de datos (BCSS, 2013).

Para evaluar el impacto sobre el medioambiente de los centros de datos hay que tener en cuenta, además del consumo energético, otros dos factores importantes. El primero de ellos es que los centros de datos están creciendo en número de servidores y, como consecuencia, en tamaño. Esto hace que ocupen, en algunos casos, grandes extensiones de terreno y generen una gran cantidad de desperdicios. Otro factor es la energía necesaria para refrigerar los servidores, que proviene fundamentalmente de combustibles fósiles como el carbón, el petróleo o el gas natural, que son muy contaminantes.

¿Cómo se están planteando empresas como Google o Facebook el tema del impacto sobre el medioambiente que causan sus centros de datos? Según el informe presentado por Greenpeace en abril de 2012 (Greenpeace, 2012), que recoge un análisis de 14 empresas TIC líderes en el sector de la computación en la nube, en general, hay signos positivos de colaboración por parte de las empresas para mejorar y utilizar diseños TIC con eficiencia energética. Además, cada vez son más las empresas que están adoptando medidas para que su demanda de energía pueda cubrirse con fuentes de energía renovable.

## **El Big Data y la ética**

Además de los daños que puede causar al medioambiente, el Big Data plantea importantes cuestiones éticas.

## La privacidad

Una primera cuestión que debe abordarse es la de quién es el dueño de los datos que se analizan. Los datos que se recogen sobre nuestras llamadas telefónicas, por ejemplo, ¿pertencen a la persona que llama, a la compañía telefónica o a cualquier agencia de espionaje del gobierno que pueda acceder a ellos? Cuando nuestros coches sean monitorizados mediante sensores, todos los datos que generen, ¿pertencerán a los conductores, a los propietarios, o a los fabricantes del vehículo? (TCS, 2013)

Tradicionalmente las organizaciones han venido utilizando diversas estrategias para asegurar la privacidad, entre ellas, anonimizar los datos de las identidades reales, creando así conjuntos de datos anónimos.

Normativas vigentes relacionadas con la privacidad, como la Ley de Protección de Datos de Carácter Personal o LOPD, recogen distintos principios relacionados con la privacidad. Uno de ellos es el principio de minimización, que establece que sólo deben guardarse aquellos datos personales que sean necesarios para conseguir objetivos legítimos y especificados, y que éstos deben ser destruidos tan pronto como no sean relevantes para la consecución de dichos objetivos (Tene y Polonetsky, 2012).

Otro principio importante para proteger la privacidad, también recogido en la LOPD, es el de “información y consentimiento” según el cual debe explicarse a los individuos qué uso se dará a la información que se recoja sobre ellos y éstos, a su vez, deben poder decidir si quieren o no que se recojan sus datos (Velásquez, 2014).

Llevar a la práctica todos estos principios, en lo que al Big Data se refiere, no parece tarea fácil. Por un lado, existen técnicas para desanonimizar o restablecer la identidad de los datos, como la que utilizaron a principios de 2013 varios expertos en seguridad informática para identificar los apellidos correspondientes a 50 genomas humanos, supuestamente anónimos (Gymrek et al., 2013).

Por otro, gran parte del valor de los datos suele estar en usos secundarios distintos de aquellos para los que se recogieron inicialmente, por lo que existe el riesgo de que los datos personales y privados se recojan y analicen con fines que los sujetos de los datos ni conocen ni aprueban. Esto hace que el mecanismo de información y consentimiento pueda no ser suficiente.

Además, los datos personales reúnen unas características que hacen especialmente difícil la tarea de establecer normativas (Rose y Kalapesi, 2012). En primer lugar, la naturaleza digital de los datos personales implica que pueden copiarse infinitamente y distribuirse globalmente, eliminando, por tanto, muchas de las fronteras comerciales que existen para los bienes materiales. Una segunda característica es que los datos no se agotan cuando se usan, se pueden reutilizar y aumentan su valor cuando se conectan con otros datos. Por último, y a diferencia de otros bienes materiales, los datos personales están íntimamente ligados a la historia e identidad del individuo.

A pesar de estas dificultades, es necesario establecer normas que (1) protejan los datos personales frente a las infracciones y abusos intencionados y no intencionados; (2) establezcan derechos, responsabilidades y permisos para que se equilibren los intereses de todos los participantes en el intercambio de datos; y (3) definan los requisitos que una empresa debe cumplir para que pueda ser considerada responsable con respecto a la protección, la seguridad y el uso de datos personales.

La privacidad no debe preocupar sólo a los gobiernos y a las empresas, sino también a los usuarios, que podemos protegerla utilizando, por ejemplo, la red TOR<sup>5</sup>, donde los mensajes intercambiados entre los usuarios no revelan su identidad porque viajan desde el origen al destino a través de ordenadores seleccionados al azar.

Existen también las credenciales anónimas que permiten a un usuario demostrar que posee autorización para acceder a un sistema sin

---

<sup>5</sup> <https://www.torproject.org/>

tener que revelar ninguna otra información sobre sí mismo, o protocolos como el de la recuperación privada de información (PIR)<sup>6</sup>, que permite a un usuario obtener información de una base de datos sin que se sepa qué información concreta se ha consultado.

## La transparencia

Además de proteger la privacidad de los datos, es necesario que los individuos tengan acceso a los datos que se recogen sobre ellos. Existen varias iniciativas en este sentido. Una de ellas es el proyecto midata del gobierno de Reino Unido (Cable, 2014) cuyos objetivos son, por un lado, conseguir que las empresas proporcionen a los consumidores acceso electrónico y seguro a los datos personales que han recogido sobre ellos y, por otro, animar a las empresas a desarrollar aplicaciones que ayuden a los consumidores a usar sus datos de forma efectiva.

Proporcionando acceso a estos datos, se consigue un mayor empoderamiento de los consumidores, porque pueden entender mejor su propio comportamiento, realizar elecciones más informadas de productos y servicios, y gestionar sus vidas de forma más eficiente. Tener acceso, por ejemplo, a los datos que una compañía de telefonía móvil tiene sobre nuestros usos y consumos puede ayudarnos a elegir una tarifa mejor.

En Estados Unidos han surgido iniciativas similares al proyecto midata, como el Blue Button<sup>7</sup>, un símbolo que aparece en determinadas páginas web y permite a los pacientes acceder a sus historiales médicos. De forma equivalente, existen las iniciativas Green Button para datos de consumo energético, y Red Button para datos académicos.

---

6 <http://crypto.stanford.edu/pir-library/>

7 <http://bluebuttondata.org/>

## **La pérdida de identidad**

Hablamos de proteger nuestros datos personales y poder acceder a ellos de forma transparente. Pero, ¿qué ocurre con las conclusiones que se obtienen al analizarlos? ¿Existen límites sobre el tipo de inferencias que se puede hacer o sobre las decisiones que se pueden tomar basadas en esas inferencias?

Ya hemos comentado que una gran parte de los datos que maneja el Big Data tiene que ver con los rastros que las personas dejamos en nuestra vida diaria. Las tiendas que frecuentamos, el importe de nuestras compras, la naturaleza de los productos que adquirimos, la frecuencia con la que repetimos nuestras compras, las palabras que introducimos en los motores de búsqueda, nuestra ubicación geográfica, las franjas horarias en las que habitualmente nos conectamos a la red, nuestras interacciones en redes sociales, los vídeos e imágenes que descargamos, las páginas que visitamos. El análisis de todos estos datos permite agrupar a las personas en distintos perfiles según su comportamiento.

Las recomendaciones de compra, las búsquedas adaptadas a nuestro perfil, o la analítica del aprendizaje pueden hacer que nuestra identidad personal quede diluida en la de un colectivo con características comunes. Incluso puede ocurrir que, algún día, las empresas sepan más sobre nosotros que nosotros mismos, y sean capaces de moldear nuestra identidad a su antojo.

## **La discriminación y el castigo anticipado**

Hemos visto cómo la capacidad del Big Data para predecir la probabilidad de que una persona tenga un determinado comportamiento en el futuro puede utilizarse para prevenir acciones terroristas, fraudes bancarios o el fracaso escolar pero, llevado al extremo, también podría utilizarse para discriminar o imponer castigos anticipados, antes de que se cometa el delito (Cukier y Mayer-Schonberger, 2013).

Un ejemplo de castigo anticipado podría ser, por ejemplo, denegar la libertad condicional a un preso si las predicciones muestran que hay una probabilidad alta de que cometa un asesinato, pero también podría aplicarse en decisiones que tengan que ver con el despido de un empleado, la intervención quirúrgica de un paciente, o incluso una solicitud de divorcio.

Puede que con estas intervenciones anticipadas se consiguiera una sociedad más segura, pero se perderían principios tan importantes como la presunción de inocencia, la capacidad de decisión y el libre albedrío, imprescindibles para que una sociedad pueda ser considerada mínimamente justa.

### **El peligro de exclusión**

Hasta ahora hemos hablado de riesgos como la pérdida de privacidad y de identidad, que pueden derivarse de nuestra participación o inclusión, quizá involuntaria, en el mundo del Big Data, pero existe también el riesgo de exclusión para aquellos cuya información no se recoge ni se analiza (Lerman, 2013).

Para ilustrar este riesgo imaginemos a una persona como Clara, una administrativa de 32 años que tiene un teléfono inteligente con el que puede acceder a herramientas como Google, Gmail, Spotify o Amazon. Clara usa Facebook para estar en contacto con sus amigos y tuitea fotos en Flickr e Instagram. Tiene un GPS en el salpicadero de su coche y en su monedero lleva tarjetas de crédito, un abono de transporte y tarjetas de cliente de los supermercados donde realiza habitualmente la compra. Con esta pequeña descripción, pueden ustedes imaginar la enorme cantidad de datos que Clara genera diariamente.

Pero en el mundo existen también personas como Daniel, que vive en un pueblo de León y trabaja como camarero en un restaurante donde le pagan en efectivo. No tiene móvil ni ordenador, y sólo usa Internet

en el ordenador que tienen en la biblioteca de su pueblo. Tampoco tiene coche ni GPS, porque apenas viaja, y en las raras ocasiones en que coge el autobús para ir a la capital, paga en efectivo.

Un mundo en el que se utilice el Big Data para configurar las políticas públicas de los gobiernos y las estrategias de mercado tendrá en cuenta los hábitos y preferencias de Clara y de personas que, como ella, generan diariamente grandes cantidades de información analizable, pero ignorará por completo las necesidades de Daniel.

A Daniel puede no importarle mucho que no se tengan en cuenta sus hábitos de consumo, pero quizá sí le preocupe perder oportunidades de empleo si deciden no abrir supermercados en su vecindario, que no se le ofrezcan descuentos donde habitualmente compra, que los nuevos productos que salgan al mercado no estén diseñados para cubrir sus necesidades, o tengan un precio que no se ajuste a su poder adquisitivo.

Si las preferencias y comportamientos de ciertos grupos no se tienen en cuenta a la hora de decidir cómo distribuir los bienes y los servicios públicos, el Big Data podría crear un nuevo tipo de invisibilidad que tendría implicaciones con respecto a las oportunidades económicas, la movilidad social y la participación.

Está claro que el Big Data puede transformar los negocios, el gobierno y la sociedad y puede ayudarnos a mejorar el mundo o, por el contrario, a promover la discriminación, la invisibilidad o el control de los ciudadanos por parte de los gobiernos.

Recordarán ustedes que al comienzo de esta lección planteábamos la disyuntiva de si las soluciones que aporta el Big Data son más numerosas que los problemas que crea. Cómo se resuelva esta disyuntiva dependerá, en gran parte, de que los Gobiernos, las empresas y los ciudadanos apostemos porque las transformaciones que el Big Data traerá consigo indudablemente conduzcan a un mundo más justo y más democrático.

Esta apuesta debe ser especialmente firme por parte de la Universidad porque tenemos la responsabilidad de formar a los que en el futuro tendrán que tomar las decisiones y aportar el máximo de sus capacidades en la construcción de ese mundo mejor.

Un mundo como el que describe la red para el Conocimiento Abierto, en el que “el conocimiento suponga poder para todos y no sólo para unos pocos, en el que los datos nos permitan elegir de forma informada cómo vivir y en el que la información y el conocimiento sean accesibles y transparentes para todos”.

## Referencias

Todas la referencias web han sido recuperadas el 29 de agosto de 2014.

1. Arcplan (2012). *Big Data FAQs. A primer*. Business Intelligence Blog. 23/03/2012. [www.arcplan.com/en/blog/2012/03./big-data-faqs-a-primer/](http://www.arcplan.com/en/blog/2012/03./big-data-faqs-a-primer/)
2. Armunanto, E. (2013). *FAST: a machine to detect people who will soon break the law*. Digital Journal. 02/04/2013. <http://www.digitaljournal.com/article/347078>
3. Arora, R. (2013). *Location-based marketing and Big Data: are marketers ready?* The smart cube. 26/03/2013. <http://www.thesmartcube.com/insights/blog/business-strategy/location-based-marketing-and-big-data-are-marketers-ready/>
4. BCSS (2013). *Big data, small footprint?* The Best Computer Science Schools. <http://www.bestcomputerscienceschools.net/big-data/>
5. BOE (2013). *Ley 19/2013, de 9 de diciembre, de transparencia, acceso a la información pública y buen gobierno*. <http://boe.es/buscar/pdf/2013/BOE-A-2013-12887-consolidado.pdf>
6. Brannock, J. (2014). *I knew you before I met you: how social media has changed the way we communicate*. The Faculty Voice. 18/03/2014. <http://imerrill.umd.edu/facultyvoice1/?p=3171>
7. Brustein, J. (2014a). *Consultant Andreas Weigend on Big Data refineries*. Bloomberg Business Week. 06/03/2014. <http://www.businessweek.com/articles/2014-03-06/consultant-andreas-weigend-on-big-data-refineries>
8. Brustein, J. (2014b). *The flu tricked Google. Can Wikipedia do better?* Bloomberg Business Week. 17/04/2014. <http://www.businessweek.com/articles/2014-04-17/the-flu-tricked-google-dot-can-wikipedia-do-better>
9. Cable, V. (2014). *Providing better information and protection for consumers*. Department for Business, Innovation & Skills. Gov.uk. <https://www.gov.uk/government/policies/providing-better-information-and-protection-for-consumers/supporting-pages/personal-data>

10. Capgemini (2014). *The Internet of Things: are organizations ready for a multi-trillion dollar prize?* Capgemini Consulting. <http://www.capgemini-consulting.com/resource-file-access/resource/pdf/the-internet-of-things.pdf>
11. Cartesian (2014). *Using mobile data for development*. Bill and Melinda Gates Foundation. <http://www.cartesian.com/wp-content/upload/Using-Mobile-Data-for-Development.pdf>
12. Cary, N. C. (2012). *Can a country's online "mood" predict unemployment spikes?* SAS. 12/03/2012. [http://www.sas.com/en\\_us/news/press-releases/2012/march/un-sma.html](http://www.sas.com/en_us/news/press-releases/2012/march/un-sma.html)
13. Cmons, M. (2012). *Sensor network to protect the elderly*. National Science Foundation. 06/12/2012. [http://www.nsf.gov/discoveries/disc\\_summ.jsp?cntn\\_id=126247&org=CISE](http://www.nsf.gov/discoveries/disc_summ.jsp?cntn_id=126247&org=CISE)
14. Clancy, H. (2013). *Jonathan Koomey: stop worrying about IT power consumption*. GreenBiz Group. 07/08/2013. <http://www.greenbiz.com/blog/2013/08/07/jonathan-koomey-stop-worrying-about-it-power-consumption>
15. CRA (2011). *Advancing discovery in science and engineering. The role of basic computing research*. Computing Research Association. [http://www.cra.org/ccc/files/docs/Natl\\_Priorities/web\\_data\\_spring.pdf](http://www.cra.org/ccc/files/docs/Natl_Priorities/web_data_spring.pdf)
16. Cramer, I. (2012). *Big Data and the Internet of Things*. Business Technology, 9. [https://www.bosch-si.com/media/bosch\\_software\\_innovations/documents/publication/english\\_1/2012/2012-07-BigData\\_IndustrialIT\\_byIMCramer\\_published\\_on\\_bosch-sicom.pdf](https://www.bosch-si.com/media/bosch_software_innovations/documents/publication/english_1/2012/2012-07-BigData_IndustrialIT_byIMCramer_published_on_bosch-sicom.pdf)
17. Cukier, K., Mayer-Schonberger, V. (2013). *Big Data: a revolution that will transform how we live, work, and think*. John Murray General Publishing Division.
18. Davenport, T.H., Dyché, J. (2013). *Big Data in big companies*. International Institute for Analytics. <http://www.sas.com/resources/asset/Big-Data-in-Big-Companies.pdf>
19. Dembosky, A. (2012). *Data prescription for better healthcare*. Financial Times. 11/12/2012. <http://www.ft.com/intl/cms/s/2/55cbca5a-4333-11e2-aa8f-00144feabdc0.html#axzz36QYbopg0>

20. Díaz, J. (2014). *Analíticas de aprendizaje: consideraciones éticas*. Aprendizaje y conocimiento: el ecosistema digital y su impacto en la formación del siglo XXI. 03/02/2014. <http://javierdisan.com/tag/big-data/>
21. Doctorow, C. (2007). *Little Brother*. <http://craphound.com/littlebrother/download/>
22. Einav, L., Levin, J. (2013). *The data revolution and economic analysis*. Stanford University. <http://web.stanford.edu/~jdlevin/Papers/BigData.pdf>
23. Ferguson, R. (2014). *Learning analytics don't just measure students' progress – they can shape it*. The Guardian. 26/04/2014. <http://www.theguardian.com/education/2014/mar/26/learning-analytics-student-progress>
24. Gartner (2012). *Gartner says EMEA IT spending will grow 1.4 percent in 2013 after declining in 2012*. Nota de prensa. 05/11/2012. <http://www.gartner.com/newsroom/id/2225616>
25. Gartner (2014). *Gartner says the Internet of Things will transform the data center*. Gartner News Room. <http://www.gartner.com/newsroom/id/2684616>
26. Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L. (2009). *Detecting influenza epidemics using search engine query data*. Nature 457, 1012-1014. <http://li.mit.edu/Stuff/CNSE/Paper/Ginsberg09Mohebbi.pdf>
27. Global Pulse (2013). *Mobile phone network data for development*. United Nations. [http://www.unglobalpulse.org/sites/default/files/Mobile%20Data%20for%20Development%20Primer\\_Oct2013.pdf](http://www.unglobalpulse.org/sites/default/files/Mobile%20Data%20for%20Development%20Primer_Oct2013.pdf)
28. Gonçalves, B., Sánchez, D. (2014). *Crowdsourcing Dialect Characterization through Twitter*. Institute for Cross-Disciplinary Physics and Complex Systems. <http://ifisc.uib-csic.es/publications/publication-detail.php?indice=2550>
29. Grant, E. (2012). *The promise of big data*. Harvard School of Public Health. HSPH News. <http://www.hsph.harvard.edu/news/magazine/spr12-big-data-tb-health-costs/>

30. Greenpeace (2011). *How dirty is your data? A Look at the energy choices that power cloud computing*. Greenpeace International. <http://www.greenpeace.org/international/Global/international/publications/climate/2011/Cool%20IT/dirty-data-report-greenpeace.pdf>
31. Greenpeace (2012). *How clean is your cloud?* Greenpeace International. <http://www.greenpeace.org/international/en/publications/Campaign-reports/Climate-Reports/How-Clean-is-Your-Cloud/>
32. Griggs, B., Kelly, H. (2013). *With Twitter going public: 23 key moments from Twitter history*. CNN. 19/09/2013. <http://www.cnn.com/2013/09/13/tech/social-media/twitter-key-moments/>
33. Gunelius, S. (2014). *The data explosion in 2014 minute by minute – Infographic*. Newstex. 12/07/2014. <http://newstex.com/2014/07/12/the-data-explosion-in-2014-minute-by-minute-infographic/>
34. Gurin, J. (2014). *Big data and open data: what's what and why does it matter?* The Guardian. 15/04/2014. <http://www.theguardian.com/public-leaders-network/2014/apr/15/big-data-open-data-transform-government>
35. Guthrie, D. (2013). *The coming Big Data education revolution*. US News. 15/08/2013. <http://www.usnews.com/opinion/articles/2013/08/15/why-big-data-not-moocs-will-revolutionize-education>
36. Gymrek, M., McGuire, A. L., Golan, D., Halperin, E., Erlich, Y. (2013). *Identifying personal genomes by surname inference*. Science, 339 (6117), 321-324. <http://www.sciencemag.org/content/339/6117/321>
37. Hood, L.E., Galas, D.J. (2008). *P4 Medicine: personalized, predictive, preventive, participatory. A Change of view that changes everything*. Computing Research Association. [http://www.cra.org/ccc/files/docs/init/P4\\_Medicine.pdf](http://www.cra.org/ccc/files/docs/init/P4_Medicine.pdf)
38. IDC (2012). *World wide Big Data technology and services 2012 – 2016 forecast*. Market Analysis. <http://laser.inf.ethz.ch/2013/material/breitman/additional%20reading/Worldwide%20>

[Big%20Data%20Technology%20and%20Services%202012-2016%20Forecast.pdf](#)

39. IDC (2014). *The digital universe of opportunities: rich data and the increasing value of the Internet of Things*. <http://www.emc.com/leadership/digital-universe/2014iview/index.htm>
40. Interxion (2013). *Big Data - beyond the hype*. [http://www.interxion.com/Documents/Whitepapers%20and%20PDFs/Big%20Data/BigData\\_Beyond\\_the\\_Hype\\_eng.pdf](http://www.interxion.com/Documents/Whitepapers%20and%20PDFs/Big%20Data/BigData_Beyond_the_Hype_eng.pdf)
41. Iron Mountain (2013). *Healthcare tackles Big Data: the tiered storage solution*. <http://www.ironmountain.com/Knowledge-Center/Reference-Library/View-by-Document-Type/General-Articles/H/Healthcare-Tackles-Big-Data-The-Tiered-Storage-Solution.aspx>
42. Jeffery, S. (2011). *Ushahidi: crowdmapping collective that exposed Kenyan election killings*. The Guardian. News Blog. 07/04/2011. <http://www.theguardian.com/news/blog/2011/apr/07/ushahidi-crowdmap-kenya-violence-hague>
43. Kalill, T. (2012). *Big Data is a big deal*. The White House. Office of Science and Technology Policy. 29/03/2012. <http://www.whitehouse.gov/blog/2012/03/29/big-data-big-deal>
44. KNC (2014). *The backlash against big data*. The Economist. 20/04/2014. <http://www.economist.com/blogs/economist-explains/2014/04/economist-explains-10>
45. Konkel, F. (2013a). *Tweets give USGS early warning on earthquakes*. FCW. The Business Federal Technology. 06/02/2013. <http://fcw.com/articles/2013/02/06/twitter-earthquake.aspx>
46. Konkel, F. (2013b). *Boston probe's big data use hints at the future*. FCW. The Business Federal Technology. 26/04/2013. <http://fcw.com/Articles/2013/04/26/big-data-boston-bomb-probe.aspx?m=1&Page=1>
47. Koomey, J. (2011). *Growth in data center electricity use 2005 to 2010*. Analytics Press. <http://www.analyticspress.com/datacenters.html>
48. Koomey, J. (2012). *The computing trend that will change everything*. MIT Technology Review. 09/04/2012. <http://www.technologyreview.com/news/427444/the-computing-trend-that-will-change-everything/>

49. Kroes, N. (2014). *Commission urges governments to embrace potential of big data*. Horizon 2020 Projects. 03/07/2014. <http://horizon2020projects.com/il-ict/commission-urges-governments-to-embrace-potential-of-big-data/>
50. Laney, D. (2001). *3D data management: controlling data volume, velocity and variety*. META Group. 06/02/2001. <http://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
51. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M. (2009). *Computational social science*. Science, 323, 721-723. <http://gking.harvard.edu/files/LazPenAda09.pdf>
52. Leonhard, G. (2014). *Big data: ¿gran negocio, gran hermano?* CNN. 26/04/2014. <http://cnnespanol.cnn.com/2014/02/26/big-data-gran-negocio-gran-hermano/>
53. Lerman, J. (2013). *Big Data and its exclusions*. Stanford Law Review. 03/09/2013. <http://www.stanfordlawreview.org/online/privacy-and-big-data/big-data-and-its-exclusions>
54. Lobo, R. (2014). *Could Songdo be the world's smartest city?* World Finance. 21/01/2014. <http://www.worldfinance.com/inward-investment/could-songdo-be-the-worlds-smartest-city>
55. Llinares, J. (2007). *Open government: la idea*. Javier Llinares, open governance, administración pública y otros temas. 29/12/2007. <http://www.javierllinares.es/?p=476>
56. MacKinnon, A. (2013). *What's so big about big data?* Online Educa Berlin News Portal. 25/09/2013. [http://www.online-educa.com/OEB\\_Newsportal/whats-so-big-about-big-data/](http://www.online-educa.com/OEB_Newsportal/whats-so-big-about-big-data/)
57. Marcus, G., Davis, E. (2014). *Eight (no, nine!) problems with Big Data*. The New York Times. 06/04/2014. [http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?\\_r=0](http://www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html?_r=0)
58. McKinsey (2011). *Big Data: the next frontier for innovation, competition, and productivity*. McKinsey Global Institute. [http://www.mckinsey.com/insights/business\\_technology/big\\_data\\_the\\_next\\_frontier\\_for\\_innovation](http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation)

59. McKinsey (2013). *Leveraging Big Data to optimize digital marketing*. McKinsey & Company. [http://www.mckinsey.com/client\\_service/marketing\\_and\\_sales/latest\\_thinking/leveraging\\_big\\_data\\_to\\_optimize\\_digital\\_marketing](http://www.mckinsey.com/client_service/marketing_and_sales/latest_thinking/leveraging_big_data_to_optimize_digital_marketing)
60. Mello, U., Treinish, L. (2012). *Precision agriculture, using predictive weather analytics to feed future generations*. IBM Research. [http://www.research.ibm.com/articles/precision\\_agriculture.shtml](http://www.research.ibm.com/articles/precision_agriculture.shtml)
61. Mills, M. P. (2013). *The cloud begins with coal. Big Data, big networks, big infrastructure, and big power. An overview of the electricity used by the global ecosystem*. Digital Power Group. [http://www.tech-pundit.com/wp-content/uploads/2013/07/Cloud\\_Begins\\_With\\_Coal.pdf](http://www.tech-pundit.com/wp-content/uploads/2013/07/Cloud_Begins_With_Coal.pdf)
62. Naughton, J. (2014). *We're all being mined for data – but who are the real winners?* The Guardian. 08/06/2014. <http://www.theguardian.com/technology/2014/jun/08/big-data-mined-real-winners-nsa-gchq-surveillance>
63. Nielson, B. (2013). *MOOC analytics: what corporate training can learn from Big Data*. Your Training Edge. 24/06/2013. <http://www.yourtrainingedge.com/mooc-analytics-what-corporate-training-can-learn-from-big-data/>
64. Normandeau, K. (2013). *Beyond volume, variety and velocity is the issue of big data veracity*. insideBIGDATA. 12/09/2013. <http://inside-bigdata.com/2013/09/12/beyond-volume-variety-velocity-issue-big-data-veracity/>
65. Noyes, D. (2014). *The top 20 valuable Facebook statistics*. Zephoria Internet Solutions. 13/06/2014. <http://zephoria.com/social-media/top-15-valuable-facebook-statistics/>
66. OpenScientist (2011). *Finalizing a definition of “citizen science” and “citizen scientists”*. 03/09/2011. <http://www.openscientist.org/2011/09/finalizing-definition-of-citizen.html>
67. Pardo Gimilio, D. (2014). *Crowdsourcing en procesos de gobierno abierto*. Universitat Politècnica de València. <http://riunet.upv.es/handle/10251/38624>
68. Pereda, C. F., Faus, J. (2014). *El Big Data llega al fútbol*. El País. 11/07/2014. [http://deportes.elpais.com/deportes/2014/07/11/mundial\\_futbol/1405101085\\_144164.html](http://deportes.elpais.com/deportes/2014/07/11/mundial_futbol/1405101085_144164.html)

69. Poynter, R. (2014). *3 tips from Big Data from Nate Silver's 'The Signal and the Noise'*. Vision Critical Blog. 15/01/2014. <http://www.visioncritical.com/blog/3-tips-big-data-nate-silver-s-signal-and-noise>
70. Prats, J. (2012). *El bajo coste llega a la secuenciación del genoma humano*. El País. 30/01/2012. [http://sociedad.elpais.com/sociedad/2012/01/30/actualidad/1327946178\\_504063.html](http://sociedad.elpais.com/sociedad/2012/01/30/actualidad/1327946178_504063.html)
71. Preciado, L. (2012). *Gobierno abierto y transparencia de la mano del "big data"*. IV Jornadas ASTICNET. ASTIC Boletic, 62, 44-45. [http://www.astic.es/sites/default/files/articulosboletic/mono9\\_laura\\_preciado.pdf](http://www.astic.es/sites/default/files/articulosboletic/mono9_laura_preciado.pdf)
72. Regalado, A. (2013). *The data made me do it. The next frontier for Big Data is the individual*. MIT Technology Review. 03/03/2013. <http://www.technologyreview.com/news/514346/the-data-made-me-do-it/>
73. Renton, S. (2014). *Snooping on students' digital footprints won't improve their experiences*. The Guardian. 26/03/2014. <http://www.theguardian.com/education/2014/mar/26/students-digital-footprints-experience>
74. Ridley, M. (2012). *Following the crowd to citizen science*. The Wall Street Journal. 11/02/2012. <http://online.wsj.com/news/articles/SB10001424052970204369404577206990249077050?mg=reno64-wsj&url=http%3A%2F%2Fonline.wsj.com%2Farticle%2FSB10001424052970204369404577206990249077050.html>
75. Rose, J., Kalapesi, C. (2012). *Rethinking personal data: strengthening trust*. BCG Perspectives. 16/05/2012. [https://www.bcgperspectives.com/content/articles/digital\\_economy\\_technology\\_software\\_rethinking\\_personal\\_data\\_strengthening\\_trust/](https://www.bcgperspectives.com/content/articles/digital_economy_technology_software_rethinking_personal_data_strengthening_trust/)
76. TCS (2013). *The emerging big returns on Big Data. A TCS 2013 global trend study*. [http://www.tcs.com/SiteCollectionDocuments/Trends\\_Study/TCS-Big-Data-Global-Trend-Study-2013.pdf](http://www.tcs.com/SiteCollectionDocuments/Trends_Study/TCS-Big-Data-Global-Trend-Study-2013.pdf)
77. Tene, O., Polonetsky, J. (2012). *Privacy in the age of Big Data. A time for big decisions*. Stanford Law Review. 02/02/2012.

- <http://www.stanfordlawreview.org/online/privacy-paradox/big-data>
78. Van Rijmenam, M. (2013a). *The industrial Internet will bring a revolution to the manufacturing industry*. Bigdata-startups. 14/10/2013. <http://www.bigdata-startups.com/industrial-internet-bring-revolution-manufacturing-industry/>
  79. Van Rijmenam, M. (2013b). *Why sentiment analytics should be a no-brainer for organisations*. Bigdata-startups. 01/10/2013. <http://www.bigdata-startups.com/sentiment-analytics-no-brainer-organisations/>
  80. Van Rijmenam, M. (2014a). *Trucking company US Xpress drives efficiency with Big Data*. Bigdata-startups. <http://www.bigdata-startups.com/BigData-startup/trucking-company-xpress-drives-efficiency-big-data/>
  81. Van Rijmenam, M. (2014b). *The Los Angeles police department is predicting and fighting crime with Big Data*. BigData-startups. <http://www.bigdata-startups.com/BigData-startup/los-angeles-police-department-predicts-fights-crime-big-data/>
  82. Velásquez, H. (2014) *Big Data en el “Universo Compliance”*. Diarijuridico.com. 12/03/2014. <http://www.diarijuridico.com/big-data-en-el-universo-compliance/>
  83. Wall, M. (2014). *Health data boom heralds new era of personalised medicine*. BBC News. 13/03/2014. <http://www.bbc.com/news/business-26133269>