

## Big Data: la próxima "gran cosa" en la gestión de la información

### [Versió catalana]

JULIO ALONSO ARÉVALO  
Grupo de investigación E-LECTRA  
Universidad de Salamanca  
[alar@usal.es](mailto:alar@usal.es)

MARTA VÁZQUEZ VÁZQUEZ  
Biblioteca Abraham Zacut  
Universidad de Salamanca  
[rulinha@usal.es](mailto:rulinha@usal.es)

Según la investigación de MGI y la Oficina de Tecnología de Negocios de McKinsey (Mannyica *et al.*, 2011), la cantidad de información que genera cualquier actividad pública o privada proporciona grandes conjuntos de datos, y el análisis de estos se ha convertido en una de las bases clave para la competencia en un futuro inmediato que sustentará las nuevas oleadas de crecimiento, de productividad, innovación y excedente del consumidor. Los líderes de todos los sectores deberán tener en cuenta las consecuencias de la gestión adecuada de esta ingente cantidad de datos, no solo orientada a aquellos que deben gestionarlos de manera directa como los propios administradores de datos, sino también con quienes tienen que tomar decisiones en las organizaciones. El aumento del volumen y el detalle de la información capturada por las empresas, el aumento de los multimedia, las redes sociales y la "Internet de las cosas" van a impulsar un crecimiento exponencial de los datos en el futuro previsible.

Los servicios como las redes sociales, la web semántica e inteligente y el comercio electrónico a menudo tienen que manejar datos a una escala demasiado grande para una base de datos tradicional. A medida que aumenta la escala y la demanda, también lo hace la complejidad. Afortunadamente, la escalabilidad y la simplicidad no son mutuamente excluyentes —en lugar de utilizar una tecnología de moda, es necesario un enfoque diferente—, ya que los sistemas que utilizan grandes cantidades de datos utilizan muchas máquinas que trabajan en paralelo para almacenar y procesar datos, que introduce retos fundamentales desconocidos para la mayoría de los desarrolladores. Big Data muestra cómo construir estos sistemas usando una arquitectura que aprovecha las ventajas de hardware agrupado junto con nuevas herramientas diseñadas específicamente para capturar y analizar datos a escala web. En él se describe la escalabilidad, para entender el enfoque de los sistemas de grandes volúmenes de datos que se pueden construir a partir de un equipo pequeño (Marz; Warren, 2012).

Casi todos los analistas consideran Big Data como una de las tendencias de futuro que deberán tener en cuenta la mayoría de las empresas e instituciones. La sociedad de las tecnologías de la información y la comunicación (TIC) propicia y requiere un diluvio universal de datos, procesarlos, entenderlos y transformarlos en decisiones de valor es el reto del análisis Big Data. Vital para las empresas cuyo activo es la información. Según estima la **International Data Corporation (IDC)** hoy los datos se incrementan un 50 % al año, o sea que se duplican cada dos años. Un informe del Foro Económico Mundial declaró que los datos constituyen una nueva clase de activo económico, como la moneda o el oro. Para la revista *Forbes*, el Big Data ha sido la principal tendencia tecnológica de los últimos años que se mantendrá los próximos años. Según la IDC, hasta 2015 su crecimiento será siete veces superior al de la media de todo el sector de las TIC. En áreas tan variadas como la ciencia y los deportes, la publicidad y la salud pública se ha producido un salto hacia el descubrimiento y la toma de decisiones a partir de los datos (López García, 2013). La tendencia ligada a Big Data también es alimentada por un mejor acceso a la información, especialmente desde que la mayoría de las empresas e instituciones han desplazado sus negocios a la nube, en lo que se ha denominado *cloud computing*, lo que facilita el acceso a estos desde cualquier tiempo y lugar mediante dispositivos móviles, e incluso que estos datos puedan ser utilizados por cualquier objeto o dispositivo electrónico en lo que se ha denominado la "Internet de las cosas".

Big Data ha dejado de estar limitado al mundo de la tecnología. Hoy en día se trata de una prioridad empresarial dada su capacidad para influir profundamente en el comercio de una economía integrada a escala global. Además de proporcionar soluciones a antiguos retos empresariales, Big Data inspira nuevas formas de transformar procesos, empresas, sectores enteros e incluso la propia sociedad. Aun así, la amplia cobertura mediática que está recibiendo no nos permite distinguir claramente el mito de la realidad: ¿qué está ocurriendo realmente? (*Analytics*, 2014). Las empresas utilizan Big Data para obtener resultados centrados en el cliente, aprovechar los datos internos y crear un mejor ecosistema de información. El análisis de todos los datos disponibles está convirtiéndose en un elemento de disrupción, así como en un factor de desintermediación que está afectando a la cadena de valor, el análisis de información en grandes volúmenes, de diversas fuentes, a gran velocidad y con una flexibilidad sin precedentes, puede suponer un factor diferencial para aquellos que decidan adoptarlo (*Big Data*, 2013).

La agencia Gartner proporciona una descripción del término en la siguiente frase: Big Data se refiere al volumen, variedad y velocidad de datos estructurados y no estructurados que se vierten a través de redes en los procesadores y dispositivos de almacenamiento, así como la conversión de dichos datos para el asesoramiento empresarial. Estos elementos se pueden dividir en tres categorías distintas: volumen, variedad y velocidad.

— Volumen (terabytes, petabytes y exabytes, eventualmente): La cantidad cada vez mayor de datos creada por los seres humanos y las máquinas está poniendo un reto a los sistemas informáticos, que están luchando para almacenar, proteger y poner a disposición toda la

información para su uso futuro.

— Variedad: Big Data es también el creciente número de tipos de datos que deben ser manejados de manera diferente a partir de simple correo electrónico, registros de datos y los registros de tarjetas de crédito. Reunidos datos para estudios científicos, registros de salud, datos financieros y multimedia: fotos, presentaciones gráficas, música, audio y vídeo.

— Velocidad: Se trata de la velocidad a la que estos datos se mueve a partir de criterios de valoración en el procesamiento y almacenamiento.

La *ciencia de datos* se refiere a un área emergente de trabajo que se ocupa de la recogida, preparación, análisis, visualización, administración y conservación de grandes colecciones de información. Aunque el término de *datos científicos* parece conectar más fuertemente con áreas tales como bases de datos y la informática, incluye muchos tipos diferentes de habilidades —incluyendo habilidades no matemáticas.

Big Data es, sin la menor duda, uno de los campos más importantes de trabajo para los profesionales de las TIC. No hay área ni sector que no esté afectado por las implicaciones que este concepto está incorporando; cambian algunas herramientas, se modifican estrategias de análisis y patrones de medida. Uno de los retos y oportunidades que tienen los profesionales de la información en este entorno es el relativo a la alfabetización sobre datos en bibliotecas universitarias y de investigación. La alfabetización informacional y la alfabetización digital en las bibliotecas han sido ampliamente discutidas y aplicadas en la literatura profesional, pero hasta hace muy poco se ha prestado poca atención a la alfabetización de datos. Sin embargo, las nuevas iniciativas de gobierno electrónico y de datos abiertas en la última década han creado datos públicos ampliamente disponibles que son de gran interés para investigadores y estudiantes. El aumento de la capacidad tecnológica para procesar gran cantidad de datos (Big Data) ofrece nuevas oportunidades tanto para el profano como para el investigador. Conocer y alfabetizar sobre estas cuestiones requiere un esfuerzo de readaptación profesional para fomentar una mentalidad sobre la importancia de estos datos y la cultura de análisis, ya que se trata de la adopción de las nuevas tecnologías, ello presenta desafíos únicos para los bibliotecarios. ¿Cómo pueden las bibliotecas desempeñar su papel en este esfuerzo de recualificación para desarrollar una "mentalidad basada en datos"?

Entre las cuestiones que se plantean está la diferencia entre ¿qué es la alfabetización de datos y en qué se diferencia de su contraparte la alfabetización informacional?, ¿qué puede aportar a un plan de estudios la alfabetización de datos en instituciones de educación superior? Este trabajo trata de abordar estas cuestiones. En la parte I se examina críticamente el concepto de *alfabetización de datos* —en qué se diferencia o se asemeja a otros tipos de alfabetizaciones y por qué es importante. En la parte II se examina la alfabetización de datos en el mundo académico, incluyendo una breve reseña bibliográfica de las prácticas de instrucción recientes en este sentido. El documento concluye con un conjunto de mejores prácticas para los bibliotecarios que desean seguir la alfabetización de datos en sus instituciones y recomendaciones para futuras investigaciones (*Big Data Now*, 2011).

El mundo científico, caracterizado por tener que manejar grandes volúmenes de datos, se ha visto muy beneficiado por Big Data Analytics. Desde aplicaciones para ciencias naturales y del cosmos, como la astronomía, la botánica y la geología, hasta funcionalidades que permiten realizar análisis pormenorizados de los casos y ofrecer tratamientos más personalizados en el ámbito de las ciencias de la salud, pasando por las distintas ciencias económicas y sociales que mayor ventaja obtienen aprovechando los beneficios aportados por estas herramientas de análisis de datos —estadística, economía o sociología, entre otras.

Como se pregunta Mario Tascón (2013), "¿Va a ser Big Data una etiqueta más que añadir a las múltiples modas que hemos ido viendo a lo largo de los últimos años en el panorama de Internet y los desarrollos digitales o es una tendencia de fondo que está afectando en su totalidad a la evolución de la Web? ¿Se trata de un verdadero reto para las empresas en los próximos años o una nueva estrategia de marketing y *vaporware* de los proveedores tecnológicos? Big Data puede llegar a ser el activo más valioso de una organización o una de sus obligaciones más costosas, todo depende de las estrategias y soluciones que se pongan en marcha a corto plazo para afrontar el ingente crecimiento del volumen, la complejidad, la diversidad y la velocidad de los datos". Como veremos, es una tendencia importante para las organizaciones y sus procesos de toma de decisiones, pero en absoluto afectará de la misma forma a todas las firmas y sectores. En España, según Big Data de IDC, cerca de un 5 % de las empresas españolas ya utiliza esta tecnología.

La *ciencia de datos* se refiere a un área emergente de trabajo que se ocupa de la recogida, preparación, análisis, visualización, administración y conservación de grandes colecciones de información. Aunque el término de *datos científicos* parece conectar más fuertemente con áreas tales como bases de datos y la informática, incluye muchos tipos diferentes de habilidades —incluyendo habilidades no matemáticas—. Para algunos, el término *ciencia de datos* evoca imágenes de los estadísticos en el laboratorio mirando fijamente parpadear las pantallas de ordenador llenas de números en desplazamiento. Nada podría estar más lejos de la verdad. Tampoco muchos de estos datos disponibles en el mundo no son solo numéricos y estructurados. En este contexto, "no estructurado" significa que los datos no están dispuestos en filas y columnas ordenadas. Si bien es cierto que las empresas, las escuelas y los gobiernos utilizan gran cantidad de información numérica —ventas de productos, promedio de calificaciones y evaluaciones fiscales son algunos ejemplos—, hay un montón de otra información en el mundo distinta de la utilizada por matemáticos y estadísticos. Así, mientras que siempre es útil tener grandes habilidades matemáticas, hay mucho por hacer en el mundo de la ciencia de datos para aquellos otros tipos de datos, como aquellos que contienen palabras, listas, fotografías, sonidos y otros tipos de información. Además, la ciencia de datos va más allá del simple análisis, ya que la ciencia de datos ofrece una gama de funciones y requiere una serie de habilidades muy diferentes (Stanton, 2013).

Ya existe alguna aplicación como **Dataverse (<http://dataverse.org/>)**, una aplicación web de código abierto desarrollada por la Universidad de Harvard que permite compartir, preservar, citar, explorar y analizar datos de investigación. El programa facilita la toma de datos y los pone a disposición de los demás, y permite replicar otros trabajos de investigación. Un repositorio Dataverse aloja varios *dataverses*. Cada *dataverse* contiene un *dataset* u otros *dataverses*, y cada conjunto de datos contiene metadatos descriptivos y archivos de datos (incluyendo la documentación y el código que acompañan a los datos). Dataverse normaliza la cita de los conjuntos de datos para que sea más fácil para los investigadores publicar sus datos y obtener un mejor reconocimiento de su trabajo. Cuando se crea un conjunto de datos en Dataverse, se genera la citación y se presenta de forma automática como un marco único de código abierto y repositorio de datos de investigación, lo que hace que los datos científicos sean lo más accesibles, reutilizables, y abiertos posible. Por lo que Big Data también supone una oportunidad, un reto y un desafío profesional para documentalistas y bibliotecarios, que como expertos en la compilación, organización, gestión y difusión de la información deberemos estar alineados con la evolución de esta propuesta, que es, sin lugar a dudas, una de las grandes líneas de desarrollo profesional, por ello deberemos conocer y adquirir las destrezas necesarias para saber cómo gestionar con eficiencia esta cantidad ingente de datos para dotarlos de valor en la sociedad de la información (Torres i Viñals, 2012).

## Bibliografía

*Analytics: el uso de Big Data en el mundo real. Cómo las empresas más innovadoras extraen valor de datos inciertos* (2014). IBM Institute for Business Value. <[http://www-05.ibm.com/services/es/gbs/consulting/pdf/EI\\_uso\\_de\\_Big\\_Data\\_en\\_el\\_mundo\\_real.pdf](http://www-05.ibm.com/services/es/gbs/consulting/pdf/EI_uso_de_Big_Data_en_el_mundo_real.pdf) ([http://www-05.ibm.com/services/es/gbs/consulting/pdf/EI\\_uso\\_de\\_Big\\_Data\\_en\\_el\\_mundo\\_real.pdf](http://www-05.ibm.com/services/es/gbs/consulting/pdf/EI_uso_de_Big_Data_en_el_mundo_real.pdf)) >. [Consulta: 30/12/2015].

*Big Data: es hora de generar valor de negocio con los datos* (2013). BBVA. <[https://www.centrodeinnovacionbbva.com/sites/default/files/bigdata\\_spanish.pdf](https://www.centrodeinnovacionbbva.com/sites/default/files/bigdata_spanish.pdf) ([http://www.centrodeinnovacionbbva.com/sites/default/files/bigdata\\_spanish.pdf](http://www.centrodeinnovacionbbva.com/sites/default/files/bigdata_spanish.pdf)) >. [Consulta: 30/12/2015].

*Big Data Now* (2011). Cambridge: O'Reilly Media. <[http://www.onmeedia.com/downloads/Big\\_Data\\_Now\\_Current\\_Perspectives\\_from\\_OReilly\\_Radar.pdf](http://www.onmeedia.com/downloads/Big_Data_Now_Current_Perspectives_from_OReilly_Radar.pdf) ([http://www.onmeedia.com/downloads/Big\\_Data\\_Now\\_Current\\_Perspectives\\_from\\_OReilly\\_Radar.pdf](http://www.onmeedia.com/downloads/Big_Data_Now_Current_Perspectives_from_OReilly_Radar.pdf)) >. [Consulta: 30/12/2015].

International Data Corporation (IDC). <<http://www.idc.com/> (<http://www.idc.com/>) >. [Consulta: 30/12/2015].

López García, David (2013). *Análisis de las posibilidades de uso de Big Data en las organizaciones*. Santander: Universidad de Cantabria. <<http://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADaS.pdf?sequence=1> (<http://repositorio.unican.es/xmlui/bitstream/handle/10902/4528/TFM%20-%20David%20L%C3%B3pez%20Garc%C3%ADaS.pdf?sequence=1>) >. [Consulta: 30/12/2015].

Mannyica, James *et al.* (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. New York: McKinsey & Company. <<http://lazowska.cs.washington.edu/escience/McKinsey.big.data.pdf> (<http://lazowska.cs.washington.edu/escience/McKinsey.big.data.pdf>) >. [Consulta: 30/12/2015].

Marz, Nathan; Warren, James (2012). *Big Data: Principles and Best Practices of Scalable Realtime Data Systems*. New York: Manning Publications. <[http://www.manning.com/marz/BD\\_meap\\_ch01.pdf](http://www.manning.com/marz/BD_meap_ch01.pdf) ([http://www.manning.com/marz/BD\\_meap\\_ch01.pdf](http://www.manning.com/marz/BD_meap_ch01.pdf)) >. [Consulta: 30/12/2015].

Stanton, Jeffrey M. (2013). *An Introduction to Data Science*. Syracuse: Syracuse University. <[https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1\\_1.pdf](https://ischool.syr.edu/media/documents/2012/3/DataScienceBook1_1.pdf)>. [Consulta: 30/12/2015].

Tascón, Mario (2013). *Big Data*. Madrid: Fundación Telefónica. <[http://www.fundaciontelefonica.com/arte\\_cultura/publicaciones-listado/pagina-item-publicaciones/?itempubli=264](http://www.fundaciontelefonica.com/arte_cultura/publicaciones-listado/pagina-item-publicaciones/?itempubli=264) ([http://www.fundaciontelefonica.com/arte\\_cultura/publicaciones-listado/pagina-item-publicaciones/?itempubli=264](http://www.fundaciontelefonica.com/arte_cultura/publicaciones-listado/pagina-item-publicaciones/?itempubli=264)) >. [Consulta: 30/12/2015].

Torres i Viñals, Jordi (2012). *Del Cloud Computing al Big Data: visión introductoria para jóvenes emprendedores*. Barcelona: UOC. <[http://www.jorditorres.org/wp-content/uploads/2012/03/Del.Cloud\\_.Computing.al\\_.Big\\_.Data\\_.JordiTorres.ES\\_.pdf](http://www.jorditorres.org/wp-content/uploads/2012/03/Del.Cloud_.Computing.al_.Big_.Data_.JordiTorres.ES_.pdf) ([http://www.jorditorres.org/wp-content/uploads/2012/03/Del.Cloud\\_.Computing.al\\_.Big\\_.Data\\_.JordiTorres.ES\\_.pdf](http://www.jorditorres.org/wp-content/uploads/2012/03/Del.Cloud_.Computing.al_.Big_.Data_.JordiTorres.ES_.pdf)) >. [Consulta: 30/12/2015].

---

### Cita recomendada

Alonso Arévalo, Julio; Vázquez Vázquez, Marta (2016). "Big Data: la próxima «gran cosa» en la gestión de la información - See more at: <http://bid.ub.edu/es/36/alonso.htm#sthash.jacFPCMg.dpuf>". *BiD: textos universitaris de biblioteconomia i documentació*, núm. 36 (juny) . <<http://bid.ub.edu/es/36/alonso.htm>>. [Consulta: 07-06-2016].

**DOI:** <http://dx.doi.org/10.1344/BiD2016.36.2> (<http://dx.doi.org/10.1344/BiD2016.36.2>)