

# ***Big Data*– Análisis informétrico de documentos indexados en *Scopus* y *Web of Science***

**Celso Martínez Musiño\***

*Artículo recibido:*  
9 de octubre de 2018  
*Artículo aceptado:*  
9 de abril de 2019

*Artículo de investigación*

## RESUMEN

El fenómeno *Big Data* es reciente, como lo demuestran las escasas publicaciones sobre el tema, lo cual incentiva esta investigación cuyos objetivos son compilar y referenciar documentos académicos incluidos en las bases de datos *Scopus* y *Web of Science* y analizar los contenidos. El método empleado es la investigación descriptiva, de primera aproximación, que consistió en la búsqueda y recuperación de información en *Scopus* y *Web of Science* en el periodo 2008-2018. Se analizaron 39 documentos, los cuales corresponden a 70 autores distribuidos en 14 títulos de revistas científicas, cuyo tipo de contribución se distribuye en 19 artículos, 10 comentarios, seis cartas al editor y cuatro reseñas.

\* Dokumenta Consultoría e Integraciones, México [cmartinez@colmex.mx](mailto:cmartinez@colmex.mx)

Otro de los resultados relevantes es que hay una alta concentración de publicaciones en *Science* y *Nature*. Los fenómenos *Big Data* y la CI son de reciente cuño y se encuentran en redefiniciones y conformaciones de dominios de estudios constantes. Encontramos un interés por las investigaciones *Big Data*; por otra parte, después de un análisis conceptual, proponemos una definición de *Big Data*.

**Palabras clave:** *Big Data*; *Scopus*; *Web of Science*; Informetría

### **Big Data Informetric Analysis of Documents Indexed in Scopus and Web of Science**

*Celso Martínez Musiño*

#### **ABSTRACT**

The Big Data phenomenon is recent, as shown by the scarce publications over this matter, a fact which incentives this research whose objective is to compile and make reference to academic documents included in the data bases *Scopus* and *Web of Science*, as well as to analyze its contents. The method used was the descriptive research of first approach, which consisted in the search and retrieval of information in *Scopus* and *Web of Science* during the period 2008-2018. 39 of the analyzed documents, corresponds to 70 authors distributed in 14 titles of scientific journals, whose type of contribution is allocated in 19 articles, 10 comments, 6 letters to the editor and 4 reviews. Another relevant result is that there exists a high concentration of publications in *Science* and *Nature*. *Big Data* and IC phenomena have been recently introduced and are found in redefinitions and conformations of constant study domains. We found an interest in *Big Data* research, yet on the other hand, after a conceptual analysis, we propose the definition of Big Data.

**Keywords:** *Big Data*; *Scopus*; *Web of Science*; Informetrics

## INTRODUCCIÓN

El fenómeno del exceso de información no es reciente, pues se publicó una gran producción de documentos impresos cuando se rediseñó la imprenta con tipos móviles, y mucho mayor número de licencias papales se distribuyeron, lo que terminó la concentración del poder eclesial y se desarrollaron otras formas de expresión religiosa. Posteriormente, con la industrialización y la concentración de la población en las ciudades, los habitantes alfabetizados requerían de datos e información expresados en los diarios, e incluso periódicos de doble tiraje (matutinos y vespertinos). Recientemente, debido a las dos grandes guerras, la ciencia y la tecnología catapultaron la medición y publicación de resultados de investigación, tanto en boletines, revistas (magazines) y *journals*, principalmente en formatos impresos, originando así una “explosión de información”.

La fácil edición y distribución de documentos electrónicos, aunado al desarrollo y comercialización de la computadora personal, detonaron en la segunda mitad del siglo XX el fenómeno denominado “sociedad de la información”, que se distingue de las otras formas de producción (primitiva, agrícola e industrial) tanto por la generación de información (documental) como por su distribución masiva, a través de internet, mediante computadoras o dispositivos móviles. Esto la convierte, además, en una fuente de ingresos por la comercialización de equipo y sus aditamentos, *software* y aquellos servicios de administración y mantenimiento de sistemas automatizados, transformando las relaciones económicas y las interacciones sociales y culturales casi a nivel global.

El mundo globalizado, el fenómeno de la sociedad de la información en transición hacia las sociedades del conocimiento, la facilidad en el acceso a dispositivos móviles y la disponibilidad de programas o aplicaciones para publicar documentos en distintos formatos (texto, imágenes, sonido, video o la combinación de estos elementos) son el caldo de cultivo para la generación masiva de datos. A la masificación en la producción de datos se le ha denominado *Big Data*, término que no sólo se circunscribe a la generación de volúmenes de datos, sino que desde principios del siglo XXI casi todas las áreas del conocimiento han empleado tiempo, espacio y dedicación para su estudio.

No obstante la facilidad para publicar y distribuir la información en internet, encontramos los siguientes problemas: a) difícil acceso a las bases de datos especializadas, con excepción de las unidades instituciones de educación superior; b) proliferación de fuentes de información y oferta de cursos de capacitación de dudoso origen o calidad; c) preponderancia de la lengua inglesa para la comunicación científica; d) proliferación de noticias falsas

(*fake news*), fenómeno que puede alcanzar, afectar o interferir la comunicación científica, y e) escasa o nula cantidad de estudios acerca de la temática de nuestra investigación. Nos proponemos como objetivos de este trabajo compilar y referenciar documentos académicos incluidos en las bases de datos *Scopus* y *Web of Science* con el fin de analizar los contenidos (títulos de documentos y temáticamente) y cuantificar los resultados obtenidos.

## LA CIENCIA DE LA INFORMACIÓN, BIG DATA E INFORMETRÍA

### ***La ciencia de la información y su objeto de estudio***

La investigación se circunscribe a los parámetros de la *ciencia de la información* (CI), área del conocimiento que se dedica al estudio del fenómeno de la información, es decir, la forma cómo ésta se produce, manifiesta, distribuye y utiliza. Sólo por mencionar una definición de *información* a partir de la CI, Faibisoff y Ely (1976) señalan que, además de contener datos, la información se encuentra constituida por ideas, símbolos o un conjunto de símbolos con un significado potencial. A partir de esta conceptualización encontramos la vinculación y la justificación de nuestra investigación, donde se observa que los datos son componente esencial de la información, y ésta a su vez objeto de análisis de la CI.

### ***Big Data y sus características***

Como fenómeno reciente del siglo XXI, para fines de nuestro estudio rescatamos algunas referencias conceptuales de *Big Data* de los mismos documentos recuperados tanto en la base de datos *Scopus* como de *Web of Science*. La mayoría de las definiciones se centran en el tamaño de los datos almacenados (Russom y Big Data Analytics, 2011: 6; Ward y Barker, 2013), o bien, las perspectivas particulares como las finanzas o las áreas de servicios, donde *Big Data* “se refiere al crecimiento de los datos y cómo se usa para optimizar los procesos comerciales, crear valor para el cliente y mitigar los riesgos” (Desouza, 2014: 10).

El tamaño de los datos es importante, pero hay otros atributos; si se añaden, además del volumen, la variedad y la velocidad entonces se constituye una definición integral (Russom y Big Data Analytics, 2011: 6) o casi total. Desouza (2014: 11) añade a las características mencionadas la viscosidad (mide la resistencia al flujo de datos), la variabilidad (calcula la tasa de cambio de flujo), la veracidad (cuenta los sesgos, el ruido, la anormalidad, etc.) y la volatilidad (indica por cuánto tiempo son válidos los datos y tiempo de almacenamiento).

### ***La informetría para el análisis de contenidos textuales***

Los estudios informétricos expresan una nueva forma de análisis de la CI para el estudio científico de los flujos de información (Wormell, 1998: 211); nosotros añadiríamos, también, el análisis de los datos en tanto que éstos son el componente principal para conformar la información. Así, el término informetría se propuso “para designar el conjunto de las actividades métricas relativas a la información, cubriendo tanto la bibliometría y la cientometría” (Egge y Rousseau, en Santos y Kobashi, 2009: 159). En la actualidad la informetría ha extendido su utilización a los estudios métricos en la web (*webmetrics*).

El término *informetría* en sí mismo “abarca el estudio de los aspectos cuantitativos de la información, independientemente de la forma en que aparezca registrada y del modo en que se genere” (Tague-Sutcliffe, 1994; Almind e Ingwersen, 1997: 405). Por otra parte, de acuerdo a su objeto de estudio, la información, es aplicable al análisis de variados aspectos tales como características de la productividad de autores, el análisis de citas, el uso de la información a partir de su demanda y uso, la obsolescencia de la literatura. Para fines de nuestro estudio justificamos y nos adherimos al concepto en tanto que “La informetría es el estudio de los aspectos cuantitativos de la información en cualquier forma [...]” (Macías-Chapula, 2001: 36). De la información seleccionamos a la información escrita, y de ésta, particularmente las palabras.

### MÉTODO

Empleamos la investigación descriptiva, de primera aproximación, cuyo procedimiento fue la búsqueda y recuperación de información en las bases de datos *Scopus* y *Web of Science*, el sustento se encontró en la ciencia de la información mediante la informetría. La configuración de búsqueda se ejecutó con los términos “big” and “data” en el campo Título de las bases mencionadas, el periodo de búsqueda fue de enero de 2007 a marzo de 2018. Posterior a la selección de los documentos se procedió al análisis, para lo cual usamos el programa Adobe Acrobat Pro DC para la extracción de los textos, el programa Microsoft Word para la depuración de los textos y Voyant-tools para el estudio cuantitativo. Enseguida nos dedicamos a la redacción de los resultados, la discusión y las conclusiones. No se incluyen en el estudio documentos indexados, pero no son cartas, artículos o reseñas sino presentaciones con diapositivas, producto de discusiones o conferencias (*webinars*).

## RESULTADOS

Para el periodo seleccionado, la cantidad de documentos recuperados y analizados fue 39, los cuales corresponden a 70 autores presentes en 14 títulos de revistas científicas, distribuidos de la siguiente manera: *Science* (18), *Nature* (8), *Neuroscience* (2), *Advances in Astronomy* (1), *Computing in Science & Engineering* (1), *ECONTENT* (1), *Information Fusion* (1), *Light: Science & Applications* (1), *Journal of Big Data* (1), *Nature biotechnology* (1), *Scientific Reports* (1), *Scientific American* (1), *Science translational medicine* (1) y *Review of Policy Research* (1). Por tipo de contribución tenemos 19 artículos, 10 comentarios, seis cartas al editor y cuatro reseñas (*Anexo*).

**Artículos**

En el periodo estudiado se localizaron 19 artículos, de los cuales seis son de autoría individual: Susan Athey (2017); Atul J. Butte, (2017); Yves Frégnac (2017); Johannes Söding (2017); Jacques Bughin (2016); Nicholas B. Turk-Browne (2013). Los 13 artículos restantes son multiautorales, en los cuales la cantidad de autores fluctúa entre dos y cinco elementos con la peculiaridad de que ningún individuo repite responsabilidad autoral. Los artículos localizados corresponden a 11 títulos de *journals* distribuidos de la siguiente manera: *Science* (8), *Nature neuroscience* (2), *Advances in Astronomy* (1), *Computing in Science & Engineering* (1), *Information Fusion* (1), *Journal of Big Data* (1), *Light: Science & Applications* (1), *Nature biotechnology* (1), *Review of Policy Research* (1), *Science translational medicine* (1) y *Scientific Reports* (1). Como se puede observar, casi la mitad de los artículos se concentran en la publicación *Science*, los demás documentos se distribuyen equitativamente en los demás *journals*.

Ahora bien, si ordenamos y relacionamos cronológicamente, desde el más reciente, los artículos con sus respectivos autores, en primer lugar encontramos a Sergio Ramírez-Gallego, Alberto Fernández, Salvador García, Min Chen y Francisco Herrera (2018), les siguen Susan Athey (2017), Atul J. Butte (2017), Yves Frégnac (2017), Johannes Söding (2017), Jacques Bughin (2016), Matthew S. Lebo, Sutti Sheila y Robert C. Green (2016), Nataša Pržulj y Noël Malod-Dognin (2016), Periklis Papakonstantinou, David P. Woodruff y Guang Yang (2016), Julian Borrill, Reijo Keskitalo y Theodore Kisner (2015), Liran Einav y Jonathan Levin (2014), Adam R. Ferguson, Jessica L. Nielson, Melissa H. Cragin, Anita E. Bandrowski y Maryann E. Martone (2014), Min Gu, Xiangping Li y Yaoyu Cao (2014), Connie L. McNeely y Jong on Hahm (2014), Muin J. Khoury y John P. A. Ioannidis (2014), David Lazer, Ryan Kennedy, Gary King y Alessandro Vespignani (2014a), Jeff W. Lichtman, Hanspeter Pfister y Nir

Shavit (2014), Nicholas B. Turk-Browne (2013), y Meyer Z. Pesenson, Isaac Z. Pesenson y Bruce McCollum (2010). En el rubro de la distribución de autores verificamos que ninguno repite autoría; la mitad de las publicaciones se registraron en el periodo 2015 y lo que va de 2018 al momento de ejecutar la búsqueda y recuperación de la información.

La suma de palabras de los títulos de los 19 artículos muestra un total de 174 palabras, pero si omitimos los artículos, los verbos, los pronombres y los nexos o preposiciones, y contando únicamente los sustantivos, adjetivos y adverbios, tenemos como resultado 114 unidades lingüísticas (palabras); si consideramos que algunas se repiten, contabilizamos únicamente 59 palabras, de las cuales *big* y *data* son las que aparecen con más frecuencia y suman 45 menciones, de allí el resto de los términos sólo contabilizan una o dos apariciones (*Anexo*).

### Comentarios

Este tipo de documentos no son una editorial, artículos o reseñas de libros, se trata de mensajes breves de especialistas que usualmente tratan la temática del número o volumen correspondiente a la revista académica en la que se incluyen. De este tipo de contribución encontramos 10 títulos: “Big Data, big picture: Metabolomics meets systems biology” (May, 2017); “The State of Big Data” (Kho, 2016); “Big Data: The power of petabytes” (Eisenstein, 2015); “Dealing with Big Data” (Osborne, 2015); “Big Data stalking. Data brokers cannot be trusted to regulate themselves” (Crawford, 2014); “Life science technologies: Big biological impacts from Big Data” (May, 2014); “The human face of Big Data” (Smolan, 2013); “Saving Big Data from Itself” (Pentland, 2014); “Big Data: Distilling meaning from data” (Frankel y Reid, 2008), y “Big Data: Data wrangling” (Goldston, 2008). En esta relación de autores comentaristas, a diferencia de los artículos, Mike May repite con dos documentos: “Big Data, big picture: Metabolomics meets systems biology” (2017) y “Life science technologies: Big biological impacts from Big Data” (2014); ambos se publicaron en *Science*.

De los 10 títulos encontrados, omitiendo los artículos, preposiciones y verbos, tenemos 26 palabras con 47 menciones (*Anexo*), entre las que predominan *Big* y *Data*; el resto de las palabras sólo reciben una mención. Por otra parte, y a diferencia de los artículos, donde encontrábamos mayormente los títulos de las revistas científicas *Nature* y *Science*, en los comentarios hay más variedad, así podemos mencionar *ECONTENT* y *Scientific American* donde publicaron Nancy Davis Kho (2016) y Kate Crawford (2014), respectivamente.

### ***Cartas al editor***

Las cartas al editor son participaciones, generalmente cortas, en las cuales se hacen aclaraciones, sugerencias o crítica a documentos ya publicados anteriormente. De este tipo de contribuciones encontramos seis cartas, escritas por Isaac Chun-Hai Fung, Zion Tsz Ho Tse y King-Wa Fu (2015); David Andre Broniatowski, Michael J. Paul y Mark Dredze, (2014); David Lazer, Ryan Kennedy, Gary King y Alessandro Vespignani (2014b); Ben Shneiderman (2014); Hubert Gijzen (2013), y Bernardo A. Huberman (2012). Estos autores publicaron en *Science* (Fung, Tse y Fu, 2015; Broniatowski, Paul y Dredze, 2014; Lazer *et al.*, 2014b; Shneiderman, 2014) y *Nature* (Gijzen, 2013; Huberman, 2012). De esta relación de autores, David Lazer, Ryan Kennedy, Gary King y Alessandro Vespignani (2014a) han publicado el artículo “The parable of Google Flu: traps in Dig Data analysis” también en *Science*. Si compilamos sólo los títulos de las cartas al editor, al igual que en los artículos y los comentarios, si omitimos artículos, verbos, pronombres y nexos, dejando únicamente los sustantivos, adjetivos y adverbios, tenemos 31 palabras; si consideramos que varias de éstas se repiten, contabilizamos solamente 18 palabras (*Anexo*).

### ***Reseñas de libros***

La menor cantidad de documentos recuperados con el registro *Big Data* en el título fueron cuatro reseñas de libros: “Big Data meets human health” (Farrington, 2016); “Big Data: The revolution is digitized” (Seife, 2015); “Big Data: Stealth control” (Aftergood, 2015), y “The Big Data debate” (Cate, 2014). El primer documento hace referencia a dos libros: *Self-Tracking* (Neft y Nafus, 2016) y *Crowdsourced Health How What You Do on the Internet Will Improve Medicine* (Yom-Tov, 2016); las reseñas de Charles Seife (2015) son también de dos títulos: *Big Data, Little Data, No Data: Scholarship in the Networked World* (Borgman, 2015) y *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior and Almost Everything Else* (Lohr, 2015). Por su parte, Steven Aftergood (2015) reseña el libro *The Black Box Society: The Secret Algorithms That Control Money and Information* (Pasquale, 2015). Finalmente, Fred H. Cate (2014) alude al libro *Privacy, Big Data, and the Public Good Frameworks for Engagement* (Lane *et al.*, 2014) (*Tabla 1*).



Títulos de artículos	Libros referenciados
"Big Data meets human health"	<i>Self-Tracking</i> <i>Crowdsourced Health How What You Do on the Internet Will Improve Medicine</i>
"Big Data: The revolution is digitized"	<i>Big Data, Little Data, No Data: Scholarship in the Networked World</i> <i>Data-ism: The Revolution Transforming Decision Making, Consumer Behavior and Almost Everything Else</i>
"Big Data: Stealth control"	<i>The Black Box Society: The Secret Algorithms That Control Money and Information</i>
"The Big Data debate"	<i>Privacy, Big Data, and the Public Good Frameworks for Engagement</i>

Tabla 1. Reseñas de libros  
Fuente: elaboración propia

La *Tabla 1* muestra que los títulos de libros referenciados no necesariamente contienen el binomio *Big Data* y la cantidad de palabras en los títulos de los libros no sólo es mayor, sino que es más diversa que aquellas palabras en los títulos de los artículos. Si sumamos las palabras de los títulos y de las reseñas de libros y les damos el mismo tratamiento que a los artículos, los comentarios y las cartas al editor, tenemos como resultado 10 palabras con 16 menciones, de las cuales una tercera parte corresponde a la dupla *Big Data*, en tanto que en las dos terceras parte restantes sólo se mencionan en una ocasión.

El resultado es contundente, en la medida en que la configuración de la búsqueda fue *Big Data* la consecuencia natural es que en todos los tipos de documentos se reflejan esos términos. En lo que se debe puntualizar, entonces, es en la presencia o ausencia temática para encontrar oportunidades investigativas. Así, es interesante descubrir que existe una revista cuyo título es precisamente *Journal of Big Data* en la cual se puede encontrar un destino para la generación de nuevos conocimientos. Por otra parte, respecto a la alta concentración de publicaciones en *Science* y *Nature*, es posible afirmar que estas revistas científicas se mantienen en la punta porque incluyen temas novedosos y actuales. Por otro lado, si en los títulos cuyas inclusiones de algún tipo de contribución (artículo, comentario, cartas al editor o reseña) se incluyen cantidades mínimas, entonces, esos títulos puedan ser, también, un destino para contribuciones (*Advances in Astronomy, Computing in Science & Engineering, ECONTENT, Information Fusion, Light: Science & Applications, Nature biotechnology, Review of Policy Research, Science translational medicine, Scientific American* y *Scientific Reports*).

Desde el punto de vista de los buscadores y las redes sociales, *Google* sólo ha sido mencionado por David Lazer, Ryan Kennedy, Gary King y Alessandro Vespignani (2014a) en el documento “The parable of Google Flu: traps in Big Data analysis” en la revista *Science*. Por otra parte, Twitter sólo ha sido atendido en el artículo “Twitter: Big Data opportunities” por David Andre Broniatowski, Michael J. Paul y Mark Dredze (2014) en *Science* y en la sección de cartas al editor repiten los autores Lazer *et al.*, (2014b) con “Twitter: Big Data opportunities—response” en el mismo título del *journal* el mismo año. Contrariamente, no se reflejan estudios en las bases de datos científicas (*Scopus* y *Web of Science*) acerca de Whatsapp, Facebook o Instagram que son ausencias notorias dada su alta participación en la creación, la edición, la reedición o la distribución de contenidos.

## DISCUSIÓN

Desde el punto de vista conceptual, tanto *Big Data* como CI son de reciente cuño y se encuentran en redefinición constante. Por un lado, Muin J. Khoury y John P. A. Ioannidis (2014: 1054) se expresan respecto a la dupla *Big Data* como el concepto, donde los datos pueden ser vinculados con la información. Desde el enfoque de la CI, los datos son los elementos clave para la conformación de la información después de un proceso de análisis. En la búsqueda de una definición de *Big Data* y de CI podemos afirmar que, en ambos casos, se encuentran en proceso de reconfiguración de sus dominios o campos de estudios (Elankavi, Kalaiprasath y Udayakumar, 2017: 560). Observamos, además, que los datos son unidades mínimas que crecen exponencialmente y son vehículo para conformar la información, en consecuencia, el fenómeno *Big Data* es sujeto de análisis por la CI y sus herramientas, como la informetría o la cientimetría (Egge y Rousseau, en Santos y Kobashi, 2009: 159), entre otras. Entonces, una vez examinadas las propuestas conceptuales, por *Big Data* replanteamos y proponemos la siguiente definición: *el fenómeno de generación masiva y constante de datos, cuyo tratamiento y organización requieren tanto de recursos tecnológicos como de programas especializados de computadora y herramientas de interpretación y análisis para lograr la científicidad.*

De acuerdo con los resultados obtenidos es posible, para futuras investigaciones, profundizar en el análisis interno de cada documento recuperado para encontrar la densidad del vocabulario o el análisis del discurso. En otro orden, con la misma relación de artículos, comentarios, cartas al editor y reseñas, proponemos indagar el factor de impacto o identificar la calidad de las revistas científicas, o bien, plantear los cuestionamientos siguientes:

¿tiene o debe tener la misma validez una reseña, los comentarios o las cartas al editor que los artículos científicos? Continuando con las derivaciones de esta investigación, se hallaron fortalezas y debilidades. Por ejemplo, hay una concentración de documentos recuperados en *Science* y *Nature* que el *Journal Citation Report* categoriza como publicaciones multidisciplinares. En este rubro multitemático, aunque con menor cantidad de escritos que *Science* y *Nature*, identificamos las revistas *Scientific Reports* y *Scientific American*, y en un sentido interdisciplinar, también con escasos artículos, podemos mencionar los *journals Big Data, Computing in Science & Engineering e Information Fusion*, en los cuales se circunscriben varios temas: ciencia, teoría y métodos computacionales o aplicaciones interdisciplinares, o la inteligencia artificial.

Por otra parte, encontramos títulos de revistas académicas con tópicos muy especializados que tratan temas como neurociencias, biotecnología o microbiología aplicada (*Nature biotechnology*), biología celular o medicina (*Science Translational Medicine*), óptica (*Light-Science & Applications*), astronomía o astrofísica (*Advances in Astronomy*), ciencia política o administración pública (*Review of Policy Research*), ciencia bibliotecaria y de la información (*ECONTENT*). Además, al localizar sólo un artículo de economía, pueden sugerirse, para posteriores investigaciones, estudios cuantitativos o bien la compilación y aplicación de herramientas para la extracción de macrodatos y sus respectivos análisis e interpretación, por ejemplo, la econometría (Varian, 2014: 5), o la combinación e integración de diferentes utensilios y modelos como sugieren Kambatla *et al.* (2014: 2568), lo cual puede redundar en nuevas oportunidades y retos (Tabares y Hernández, 2014: 1).

La CI y estudios de *Big Data* identifican, respectivamente, a la información y los datos como elementos u objetos de estudio, pero hay voces, como las de Silvia Martínez-Martínez y Pablo Lara-Navarra (2014), que sugieren analizar los medios sociales, con lo cual estamos de acuerdo. Finalmente, podemos formular otras propuestas investigativas y cuestionamientos. En primer lugar, proponemos considerar la demografía y plantear las relaciones intergeneracionales, la migración o los estudios de género, por mencionar algunos subtemas; en segundo término, podemos formular las siguientes preguntas: ¿cuál es el documento científico más citado?, ¿cuál será el porvenir y por cuánto tiempo se seguirá investigando el fenómeno *Big Data*? ¿Dicha dupla se consolidará como ciencia de los datos o sólo es una moda pasajera? ¿Se establecerá una confrontación, desplazamiento o complemento, tanto de la CI y la ciencia de los datos?

## CONCLUSIONES

Los objetivos de compilar y referenciar documentos académicos incluidos en las bases de datos *Scopus* y *Web of Science* y analizar los contenidos (títulos de documentos y temáticamente) se han cumplido. Podemos entonces concluir lo siguiente: *Big Data* es un fenómeno reciente, la publicación y la inclusión de documentos de distinto tipo (artículos, cartas al editor y reseñas de libros) indexados en bases de datos de alto impacto destinadas para las comunidades científicas demuestran que hay interés por la temática de esta investigación.

Los tópicos analizados muestran que hay una intersección de la temática *Big Data* con otras materias, en un grado mayor se opta por aquellas revistas académicas del tipo multi o interdisciplinar, y en mucho menor medida se eligen las publicaciones de carácter especializado, por ejemplo, las ciencias computacionales, la economía, la astronomía, la astrofísica y la óptica, por mencionar algunas. La diferenciación, la concentración de ciertas publicaciones o la ausencia de publicaciones en *journals* especializados nos permiten valorar los vacíos y por lo tanto detectar las oportunidades investigativas o destinos para publicar.

Los hallazgos encontrados nos muestran que el idioma inglés y la revista científica son los vehículos de la comunicación científica. La inclusión, la clasificación para la asignación de puntajes (factor de impacto) de variadas secciones en este tipo de publicaciones pueden abrir un debate, no es la misma inversión en un artículo que una reseña de libros. Otro dilema puede presentarse en relación al acceso a las bases de datos, aquellas de tipo cerrado con otras de acceso abierto.

## REFERENCIAS

- Aftergood, Steven. 2015. "Big Data: Stealth control". *Nature* 517 (7535): 435.
- Almind, Tomas C. y Peter Ingwersen. 1997. "Informetric analyses on the world wide web: methodological approaches to 'webometrics'". *Journal of documentation* 53 (4): 404-426.
- Athey, Susan. 2017. "Beyond prediction: Using Big Data for policy problems". *Science* 355 (6324): 483-485.
- Borgman, Christine L. 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA; London, England: MIT Press.
- Borrill, Julian, Reijo Keskitalo y Theodore Kisner. 2015. "Big bang, big data, big iron: fifteen years of cosmic microwave background data analysis at NERSC". *Computing in Science & Engineering* 17 (3): 22-29.

- Broniatowski, David Andre, Michael J. Paul y Mark Dredze. 2014. "Twitter: Big Data opportunities". *Science* 345 (6193): 148-148.
- Bughin, Jacques. 2016. "Big Data, Big bang?" *Journal of Big Data* 3 (1): 2.
- Butte, Atul J. 2017. "Big Data opens a window onto wellness". *Nature biotechnology* 35 (8): 720.
- Cate, Fred H. 2014. "The Big Data debate". *Science* 346 (6211): 818.
- Crawford, Kate. 2014. "Big Data stalking. Data brokers cannot be trusted to regulate themselves". *Scientific American* 310 (4): 14-14.
- Desouza, Kevin. 2014. *Realizing the promise of Big Data*. Washington, DC: IBM Center for the Business of Government.
- Einav, Liran y Jonathan Levin. 2014. "Economics in the age of Big Data". *Science* 346 (6210): 1243089.
- Eisenstein, Michael. 2015. "Big Data: The power of petabytes". *Nature* 527 (7576): S2.
- Elankavi, R., R. Kalaiprasath y R. Udayakumar. 2017. "Data Mining with Big Data Revolution Hybrid". *International Journal on Smart Sensing & Intelligent Systems* 10: 560-573.
- Faibisoff, Sylvia G. y Donald P. Ely. 1976. "Information and information needs". *Information Reports and Bibliographies* 5 (5): 2-16.
- Farrington, Conor. 2016. "Big Data meets human health". *Science* 353 (6296): 227.
- Ferguson, Adam R., Jessica L. Nielson, Melissa H. Cragin, Anita E. Bandrowski y Maryann E. Martone. 2014. "Big Data from small data: data-sharing in the 'long tail' of neuroscience". *Nature neuroscience* 17 (11): 1442.
- Frankel, Felice y Rosalind Reid. 2008. "Big Data: Distilling meaning from data". *Nature* 455 (7209): 30.
- Frégnac, Yves. 2017. "Big Data and the industrialization of neuroscience: A safe roadmap for understanding the brain?" *Science* 358 (6362): 470-477.
- Fung, Isaac Chun-Hai, Zion Tsz Ho Tse y King-Wa Fu. 2015. "Converting Big Data into public health". *Science* 347 (6222): 620-620.
- Gijzen, Hubert. 2013. "Development: Big Data for a sustainable future". *Nature* 502 (7469): 38.
- Goldston, David. 2008. "Big Data: Data wrangling". *Nature News* 455 (7209): 15-15.
- Gu, Min, Xiangping Li y Yaoyu Cao. 2014. "Optical storage arrays: a perspective for future Big Data storage". *Light: Science & Applications* 3 (5): e177.
- Huberman, Bernardo A. 2012. "Sociology of science: Big Data deserve a bigger audience". *Nature* 482 (7385): 308.
- Kambatla, Karthik, Giorgos Kollias, Vipin Kumar y Ananth Grama. 2014. "Trends in big data analytics". *Journal of Parallel and Distributed Computing* 74 (7): 2561-2573.
- Kho, Nancy Davis. 2016. "The State of Big Data". *ECONTENT* 39 (1): 28-29.
- Khoury, Muin J. y John P.A. Ioannidis. 2014. "Big Data meets public health". *Science* 346 (6213): 1054-1055.
- Lane, Julia, Victoria Stodden, Stefan Bender y Helen Nissenbaum, eds. 2014. *Privacy, Big Data, and the Public Good Frameworks for Engagement*. Cambridge, MA: Cambridge University Press.
- Lazer, David, Ryan Kennedy, Gary King y Alessandro Vespignani. 2014a. "The parable of Google Flu: traps in Big Data analysis". *Science* 343 (6176): 1203-1205.

- Lazer, David, Ryan Kennedy, Gary King y Alessandro Vespignani. 2014b. "Twitter: Big Data opportunities—response". *Science* 345 (6193): 148-149.
- Lebo, Matthew S., Sheila Sutti y Robert C. Green. 2016. "Big Data gets personal". *Science Translational Medicine* 8 (322): 322fs3-322fs3.
- Lichtman, Jeff W., Hanspeter Pfister y Nir Shavit. 2014. "The Big Data challenges of connectomics". *Nature Neuroscience* 17 (11): 1448.
- Lohr, Steve. 2015. *Data-ism: The Revolution Transforming Decision Making, Consumer Behavior and Almost Everything Else*. New York, NY: Harper Business.
- Macías-Chapula, César A. 2001. "Papel de la informetría y de la cienciometría y su perspectiva nacional e internacional". *Acimed* 9: 35-41.
- Martínez-Martínez, Silvia y Pablo Lara-Navarra. 2014. "El Big Data transforma la interpretación de los medios sociales". *El Profesional de la Información* 23 (6): 575-581.
- May, Mike. 2014. "Life science technologies: Big biological impacts from Big Data". *Science* 344 (6189): 1298-1300.
- May, Mike. 2017. "Big Data, big picture: Metabolomics meets systems biology". *Science* 356 (6338): 646-648.
- McNeely, Connie L. y Jong on Hahm. 2014. "The Big (Data) bang: Policy, prospects, and challenges". *Review of Policy Research* 31 (4): 304-310.
- Neft, Gina y Dawn Nafus. 2016. *Self-Tracking*. Cambridge, MA: MIT Press.
- Osborne, Ian S. 2015. "Dealing with Big Data". *Science* 349 (6247): 491-492.
- Papakonstantinou, Periklis A., David P. Woodruff y Guang Yang. 2016. "True Randomness from Big Data". *Scientific Reports* 6: 33740.
- Pasquale, Frank. 2015. *The Black Box Society: The Secret Algorithms That Control Money and Information*. Cambridge, MA; London, England: Harvard University Press.
- Pentland, Alex. 2014. "Saving Big Data from Itself". *Nature* 314 (2): 64-67.
- Pesenson, Meyer Z., Isaac Z. Pesenson y Bruce McCollum. 2010. "The data big bang and the expanding digital universe: High-dimensional, complex and massive data sets in an inflationary epoch". *Advances in Astronomy*, Article ID 350891, 1-16.
- Pržulj, Nataša y Noël Malod-Dognin. 2016. "Network analytics in the age of Big Data". *Science* 353 (6295): 123-124.
- Ramírez-Gallego, Sergio, Alberto Fernández, Salvador García, Min Chen y Francisco Herrera. 2018. "Big Data: Tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce". *Information Fusion* 42: 51-61.
- Russom, Philip y Big Data Analytics. 2011. TDWI best practices report. *Fourth Quarter* 19 (4): 1-34.
- Santos, Raimundo Nonato Macedo dos y Nair Yumiko Kobashi. 2009. "Bibliometria, cientometria, infometria: conceitos e aplicações". *Pesquisa Brasileira em Ciência da Informação* (Brasília) 2 (1): 155-172.
- Seife, Charles. 2015. "Big Data: The revolution is digitized". *Nature* 518 (7540): 480.
- Shneiderman, Ben. 2014. "The big picture for Big Data: Visualization". *Science* 343 (6172): 730-730.
- Smolan, Rick. 2013. "The human face of Big Data". *Science* 351 (6274): 673.
- Söding, Johannes. 2017. "Big-Data approaches to protein structure prediction". *Science* 355 (6322): 248-249.
- Tabares, Luis F. y Jhonatan. F. Hernández. 2014. *Big Data Analytics: Oportunidades, Retos y Tendencias*.  
[https://www.academia.edu/15043636/Big\\_Data\\_Analytics\\_Oportunidades\\_Retos\\_y\\_Tendencias](https://www.academia.edu/15043636/Big_Data_Analytics_Oportunidades_Retos_y_Tendencias)

- Tague-Sutcliffe, Jean. 1994. "Introducción a la informetría". *Acimed* 2 (3): 26-35.
- Turk-Browne, Nicholas B. 2013. "Functional interactions as Big Data in the human brain". *Science* 342 (6158): 580-584.
- Varian, Hal R. 2014. "Big data: New tricks for econometrics". *Journal of Economic Perspectives* 28 (2): 3-28.
- Ward, Jonathan S. y Adam Barker. 2013. "Undefined by data: a survey of big data definitions". Preprint, presentado el 20 de septiembre de 2013.  
<https://arxiv.org/abs/1309.5821>
- Wormell, Irene. 1998. "Informetría: explorando bases de datos como instrumentos de análise". *Ciência da Informação* 27 (2): 210-216.
- Yom-Tov, Elad. 2016. *Crowdsourced Health How What You Do on the Internet Will Improve Medicine*. Cambridge, MA; London, England: MIT Press.

*Para citar este texto:*

- Martínez Musiño, Celso. 2020. "Big Data-Análisis informétrico de documentos indexados en Scopus y Web of Science". *Investigación Bibliotecológica: archivonomía, bibliotecología e información* 34 (82): 87-102.  
<http://dx.doi.org/10.22201/iibi.24488321xe.2020.82.58035>

## Anexo

### Palabras de tipo de documento y su frecuencia de aparición en los títulos

Artículos	Frecuencia	Comentarios	Frecuencia	Cartas al editor	Frecuencia	Reseñas	Frecuencia
data	23	data	12	big	7	big	4
big	22	big	11	data	6	data	4
bang	4	biological	1	opportunities	2	control	1
age	2	biology	1	twitter	2	debate	1
analysis	2	brokers	1	audience	1	digitized	1
analytics	2	dealing	1	bigger	1	health	1
brain	2	distilling	1	converting	1	human	1
challenges	2	face	1	deserve	1	meets	1
neuroscience	2	human	1	development	1	revolution	1
policy	2	impact	1	future	1	stealth	1
prediction	2	life	1	health	1	<b>SUBTOTAL</b>	<b>16</b>
storage	2	meaning	1	picture	1		
algorithms	1	meets	1	public	1		
approaches	1	metabolomics	1	response	1		
arrays	1	petabytes	1	science	1		
background	1	picture	1	sociology	1		
beyond	1	power	1	sustainable	1		
complex	1	regulate	1	Visualization	1		
connectomics	1	saving	1	<b>SUBTOTAL</b>	<b>31</b>		
cosmic	1	science	1				
digital	1	stalking	1				
dimensional	1	state	1				
economics	1	systems	1				
epoch	1	technologies	1				
expanding	1	trusted	1				
flu	1	wrangling	1				
functional	1	<b>SUBTOTAL</b>	<b>47</b>				
fusion	1						
future	1						
gets	1						
google	1						
guidelines	1						
health	1						
high	1						
human	1						
industrialization	1						
inflationary	1						
information	1						
interactions	1						
iron	1						
mapreduce	1						
massive	1						
meets	1						
microwave	1						
nersc	1						
network	1						
opens	1						
optical	1						
parable	1						
personal	1						
perspective	1						
problems	1						
process	1						
prospects	1						
protein	1						
public	1						
randomness	1						
roadmap	1						
safe	1						
<b>SUBTOTAL</b>	<b>114</b>					<b>TOTAL</b>	<b>170</b>