



CIECTI

Centro Interdisciplinario
de Estudios en Ciencia,
Tecnología e Innovación

DT
3

BIG DATA. AVANCES RECIENTES A NIVEL INTERNACIONAL Y PERSPECTIVAS PARA EL DESARROLLO LOCAL

Facundo Malvicino y Gabriel Yoguel

BIG DATA. AVANCES RECIENTES A NIVEL INTERNACIONAL Y PERSPECTIVAS PARA EL DESARROLLO LOCAL

DOCUMENTO DE TRABAJO N° 3

Facundo Malvicino y Gabriel Yoguel

CENTRO INTERDISCIPLINARIO DE ESTUDIOS EN CIENCIA, TECNOLOGÍA E INNOVACIÓN





CIECTI

Centro Interdisciplinario
de Estudios en Ciencia,
Tecnología e Innovación

Malvicino, Facundo

Big data : avances recientes a nivel internacional y perspectivas para el desarrollo local / Facundo Malvicino ; Gabriel Yoguel. - 1a ed. - Ciudad Autónoma de Buenos Aires : CIECTI, 2016.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-987-4193-05-6

1. Bases de Datos. 2. Análisis de Datos. 3. Procesamiento de Datos. I. Yoguel, Gabriel II. Título
CDD 332

La investigación que dio base a este estudio finalizó en marzo de 2015.

©2016 CIECTI

Queda hecho el depósito que marca la ley 11.723.

Se autoriza la reproducción total o parcial de esta obra, para fines educativos u otros fines no comerciales, siempre que se cite la fuente.

Godoy Cruz 2390 – PB (C1425FQD), CABA

(54-11) 4899-5500, int.5684

www.ciecti.org.ar / info@ciecti.org.ar

Seguinos en  @ciecti

Buscanos en  /ciecti

AUTORIDADES

Presidente

Aldo Ferrer

Directora general

Ruth Ladenheim

EQUIPO EDITORIAL

Coordinación editorial

Fernando Porta

Apoyo a la coordinación

Paula Isaak y Julia Pena

Equipo de investigación

Facundo Malvicino y Gabriel Yoguel

Edición

Mara Sessa y Patricia Ferrante

Diseño gráfico

Lea Ágreda

SIGLAS

AACREA	Asociación Argentina de Consorcios Regionales de Experimentación Agrícola
CESSI	Cámara de Empresas de Software y Servicios Informáticos
OCDE	Organización para la Cooperación y el Desarrollo Económico
PBI	producto bruto interno
SSI	Software y Servicios Informáticos
SUBE	Sistema Único de Boleto Electrónico
TI	tecnologías de la información
TIC	tecnologías de la información y la comunicación

ÍNDICE

Resumen	6
Resumo / Abstract.....	7
Introducción	9
Big Data: ¿nuevo paradigma o nueva fase del paradigma intensivo en información y comunicación?	12
Impacto económico: cambios en la productividad y competitividad.....	12
Consideraciones metodológicas y epistemológicas a partir de Big Data	18
Definición de Big Data y aspectos técnicos.....	24
Análisis de Big Data. Un esquema conceptual.....	26
Big Data: aplicaciones sociales y económicas. Oportunidades para el desarrollo	28
Salud	28
Administración pública	30
Recursos naturales, medio ambiente y energía	31
Industria manufacturera	32
Transporte.....	33
Otras alternativas que ofrece Big Data	33
Conclusiones.....	35
Apéndice. Técnicas y tecnologías vigentes para el tratamiento de Big Data.....	37
Bibliografía	41

RESUMEN

Palabras clave

Big Data
TIC
desarrollo
sectores productivos

Big Data abre la posibilidad de hacer grandes avances en muchas disciplinas científicas, aportes a la innovación y mejoras de productividad y competitividad. Sin embargo, hay que tener en cuenta que los desafíos que se presentan con Big Data no solo están referidos al volumen de datos, sino también a su heterogeneidad, a su carácter poco estructurado, a los aspectos de privacidad, a los problemas de procesamiento, a la visualización, entre otros. En términos esquemáticos, la problemática Big Data puede dividirse entre los desafíos tecnológicos de infraestructura y servicios—software y hardware—, por un lado, y la dimensión del análisis, por el otro. Asimismo, junto a los problemas de infraestructura y de servicios, el aprovechamiento de Big Data para el desarrollo social y económico requiere un enfoque multidisciplinario y el diseño de políticas estratégicas.

RESUMO

Big Data abre a possibilidade de fazer grandes avanços em muitas disciplinas científicas, contribuir para a inovação e melhorias na produtividade e competitividade. No entanto, os desafios colocados pela Big Data, não são apenas ao volume de dados, mas também à sua heterogeneidade, ao seu caráter pouco estruturado, às questões de privacidade, aos problemas de processamento, à visualização, entre outros. Em termos esquemáticos, a

problemática Big Data pode ser dividida entre os desafios tecnológicos de infraestrutura e serviços –software e hardware–, por um lado, e a dimensão da análise, por outro. Além disso, junto aos problemas de infraestrutura e de serviços, o uso de Big Data para o desenvolvimento social e econômico requer uma abordagem multidisciplinar e desenho de políticas estratégicas.

Palavras-chave

Big Data
TIC
desenvolvimento
setores produtivos

ABSTRACT

Big Data is the opportunity to contribute in many scientific disciplines, innovation, productivity and competitiveness. However, the challenges posed by Big Data not only refer to the volume of data, but also their heterogeneity, privacy issues, processing and visualization, among others. Briefly, the Big Data problem can be divided within the

technological infrastructure and services challenges –software and hardware– on the one hand, and the dimension of the analysis on the other hand. Despite the infrastructure and services problems, the use of Big Data for social and economic development requires a multidisciplinary approach and the design of strategic policies.

Keywords

Big Data
ICT
development
productive sectors

Introducción

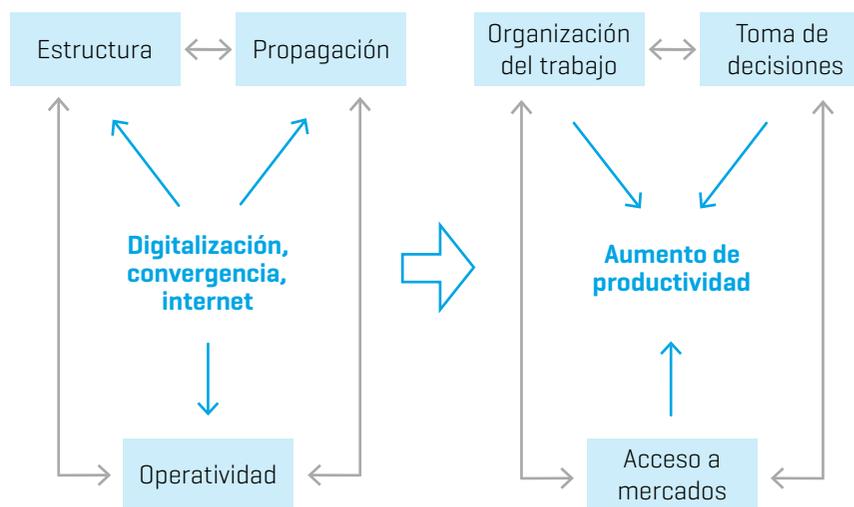
Big Data¹ surge como una nueva fase del paradigma intensivo en información y comunicación que abarca no solo su dimensión tecnológica, sino también una dimensión social, económica, política y cultural. Por lo tanto, una comprensión holística sobre el tema requiere un abordaje multidisciplinario. Al igual que con el surgimiento de las tecnologías de la información y la comunicación (TIC), existe una discusión acerca de si nos encontramos frente a un cambio de paradigma tecnológico, una nueva revolución industrial-tecnológica o si Big Data no ha hecho más que intensificar la competencia en mercados ya existentes, sin mostrar un impacto en las mejoras de la productividad general. Los autores que tienen una visión optimista sostienen que Big Data estaría provocando cambios en la organización del trabajo con mejoras en la toma de decisiones y el acceso a los mercados. Estos cambios darían lugar a una reducción en las brechas de productividad a partir de una mayor competitividad de las unidades productivas con un impacto positivo sobre la calidad de vida de las personas (McKinsey, 2011; Brynjolfsson, Hitt y Kim, 2011; UN Global Pulse, 2012; Chen y Zhang, 2014). La viabilidad de este proceso está sujeta a los cambios comunicacionales e informacionales que promueve la digitalización, la convergencia tecnológica e internet a distintos niveles: a nivel operativo se trata del acceso remoto y en tiempo real a la información para la toma de decisiones, lo que permite un desanclaje territorial; los cambios estructurales tienen que ver con la mayor y mejor accesibilidad comunicacional (conectividad); mientras que nuevos niveles de propagación permiten variaciones en la escala comunicacional (Forte *et al.*, 2012). En la figura 1 presentamos un esquema del mecanismo de transmisión entre el cambio comunicacional y el aumento de la productividad.

Existe, por lo tanto, una oportunidad para crear un sendero de desarrollo virtuoso que permita reducir las brechas de productividad con los países desarrollados, a partir del diseño de políticas efectivas orientadas a sectores de alto contenido tecnológico e informacional, en conjunto con el fortalecimiento de organismos públicos de ciencia y tecnología. Sin embargo, como veremos a lo largo del trabajo, Big Data no es de por sí una panacea, ya que requiere importantes desafíos para países en desarrollo.

1. Se utiliza el término en inglés, pues se trata de un concepto de difusión global en ese idioma.

En las últimas décadas, la Argentina ha desarrollado el sector de Software y Servicios Informáticos (SSI), que se posicionó entre los más dinámicos de la región. Por lo tanto, existen capacidades técnicas reales y potenciales que permitirían aprovechar este avance informacional-tecnológico a los fines del crecimiento y desarrollo socioeconómico del país. Entre 2003 y 2010 el sector de servicios de informática e información ha mostrado incrementos significativos en el empleo, las ventas y las exportaciones, aunque en el marco de una gran heterogeneidad (Barletta *et al.*, 2013). Según la Cámara de Empresas de Software y Servicios Informáticos (CESSI, 2014), entre 2003 y 2012 la facturación del mercado de TIC en la Argentina creció el 19,8% anual en promedio. Dadas estas características, el país cuenta con la oportunidad de crear una plataforma para una industria de Big Data mediante la promoción de políticas convergentes hacia un uso intensivo de datos digitales y TIC.

Figura 1 Mecanismo de transmisión: cambio comunicacional y productividad



Fuente: Elaboración propia.

Sin embargo, de acuerdo con los resultados de la Encuesta Big Data 2014, el mercado de este sector en la Argentina muestra una demanda escasa y de baja complejidad, explicada en parte por la falta de conocimiento sobre el tema, problemas institucionales y limitaciones en la infraestructura. Simultáneamente, las instituciones científicas revelan una desconexión entre la producción académica y los desarrollos de las empresas locales. En ese marco, el objetivo del presente documento es discutir el rol del sector público en el desarrollo de nuevos sectores de alta tecnología en países en desarrollo, tomando la situación actual de Big Data en la Argentina.

Frente a este desafío surgen las siguientes preguntas: ¿se encuentra el país en condiciones de aprovechar los beneficios que promete esta nueva fase de sociedades basadas en el conocimiento? Si el análisis de grandes paquetes de datos requiere de una base tecnológico-comunicacional, ¿está el país apto para dar respuesta a esta exigencia? ¿Cómo impacta el escaso desarrollo local de infraestructura de alta complejidad para el desarrollo de Big Data?

En ese contexto, este documento de trabajo pretende establecer un marco teórico-conceptual para desarrollar la discusión sobre los potenciales beneficios de Big Data en la Argentina. En consecuencia, interesa alcanzar una definición técnica de lo que se suele denominar Big Data para delimitar su alcance. Adicionalmente, se avanzará sobre la conceptualización del análisis de Big Data como forma ineludible de obtener un valor de esos beneficios y se hará un relevamiento de algunas de las aplicaciones sociales y económicas que se vienen desarrollando. En la primera sección se discute la dimensión epistemológica del fenómeno Big Data. En la segunda, proponemos una definición de Big Data a partir de los puntos comunes que existen en la literatura y presentamos el esquema conceptual de lo que se entiende por análisis de datos. Luego, presentamos algunas de las aplicaciones sociales y económicas orientadas a las políticas de desarrollo. Finalmente, discutimos las principales conclusiones.

Big Data: ¿nuevo paradigma o nueva fase del paradigma intensivo en información y comunicación?

Impacto económico: cambios en la productividad y competitividad

Big Data constituye un fenómeno global que puede llegar a tener un impacto económico real y potencial, que beneficie tanto al sector público como al privado en el aumento de la productividad, la competitividad sectorial y la calidad de vida de la ciudadanía. Sin embargo, existe una importante discusión acerca de si nos encontramos frente a un cambio de paradigma tecnológico, una nueva revolución industrial-tecnológica o si Big Data no solo no es un cambio de paradigma, sino que no ha hecho más que intensificar la competencia en mercados ya existentes, sin mostrar un impacto en las mejoras de la productividad general. Se pueden identificar tres grupos dentro de la literatura: los autores optimistas, que ven en Big Data un cambio positivo en la sociedad, a partir de mejoras en la productividad, en las respuestas a problemas sociales en áreas como salud y en la creación de nuevas oportunidades de negocios (McKinsey, 2011; Brynjolfsson, Hitt y Kim, 2011; UN Global Pulse, 2012; Chen y Zhang, 2014); los escépticos/pesimistas, que consideran que Big Data es una moda pasajera, que no reviste las virtudes adjudicadas o que incluso puede ser una forma de avasallamiento a las libertades individuales y la privacidad (Campanario, 2013; *The New York Times*, 2013; Gordon, 2014);² y los críticos, quienes ven oportunidades aunque no dejan de realizar un análisis crítico de este proceso, más allá de los indicadores socioeconómicos tradicionales (Peres y Hilbert, 2010; Manovich, 2012; boyd³ y Crawford, 2012).

De acuerdo con McKinsey, consultora de *management* estratégico reconocida mundialmente, nos encontramos frente a un punto de inflexión: “en la cúspide de una tremenda ola de innovación, productividad y crecimiento, como también de nuevas formas de competencia y apropiación de valor” (2011, p. 2). Si bien los datos siempre han sido parte del impacto de las TIC, los cambios que Big Data provoca impactan en el panorama económico, lo que genera nuevas oportunidades de negocios y mejoras en las tomas de decisiones a partir de

² Joel Waldfogel, de la Universidad de Minnesota, considera que las oportunidades de negocio que Big Data ha generado no son más que el canibalismo de los negocios que ya existían *offline*.

³ danah boyd utiliza su propio nombre con iniciales en minúscula.

la disponibilidad de datos en tiempo real que permiten cambios en la productividad. Brynjolfsson, Hitt y Kim (2011) realizaron entrevistas en 179 grandes firmas en Estados Unidos en 2008, y concluyeron que aquellas que incorporaron el uso de Big Data para la toma de decisiones y la creación de nuevos productos y servicios, alcanzaban una productividad entre 5% y 6% superior a la esperada, considerando sus otras inversiones y usos tecnológicos. En cuanto a capacidad de almacenamiento y generación de datos, un estudio de McKinsey (2011) estimó que en 2010 el 60% de la población mundial utilizaba celulares—cuyo 12% de los celulares eran *smartphones*— y que esta penetración crecía el 20% anual, mientras que se calcularon más de 30 millones de sensores conectados en red en el sector transporte, en automóviles, industrias, servicios públicos y comercios minoristas con una tasa de crecimiento del 30% anual. Asimismo, el estudio concluyó que a nivel global en 2010, las empresas habrían tenido capacidad de almacenar más de 7 exabytes⁴ de nuevos datos, mientras que personas particulares almacenaron más de 6 exabytes de nuevos datos en computadoras personales (PC) y *notebooks*. Es físicamente imposible almacenar todos los datos que generamos. Para tener una noción de las dimensiones, 1 exabyte equivale a 4 mil veces la información almacenada en la biblioteca del Congreso de Estados Unidos.

Estos importantes avances tienen como antecedente los cambios significativos en la capacidad de almacenamiento, procesamiento de datos y ancho de banda que se manifiestan en las leyes de Moore, Gilder y Metcalfe.⁵ De acuerdo con la Organización para la Cooperación y el Desarrollo Económico (OCDE, 2013), el incremento exponencial de generación, recolección y transporte de datos es producto del desarrollo tecnológico comunicacional—banda ancha, *smartphones*, *smart grid*, etc.— y su mayor disponibilidad, dado que el costo de acceso a internet ha venido decreciendo considerablemente en los países desarrollados.⁶ Asimismo, el uso—almacenamiento y procesamiento— de datos se vio favorecido por el descenso en el costo del almacenamiento, procesamiento y análisis: el costo de almacenamiento en unidades de discos duros (*hard disk drives*) cayó en un promedio anual del 39% entre 1998 y 2012;⁷ el costo de dispositivos de estado sólido (*solid state drives*) cayó el 51% anual entre 2007 y 2012; mientras que el costo del análisis genómico se desplomó en un promedio anual del 60% entre 2001 y 2012, y su caída se aceleró a partir de mediados de 2008.⁸ A esto debemos agregar que el aumento del acervo de información no ha sido acompañado por un avance equivalente en la capacidad de

⁴ Exabyte = 10^3 petabytes = 10^6 terabytes = 10^9 gigabytes = 10^{12} megabytes = 10^{15} kilobytes = 10^{18} bytes.

⁵ La Ley de Moore establece que el número de transistores por procesador se duplica cada 18 meses—hecho que se ha registrado en los últimos 35 años aproximadamente—; la Ley de Gilder plantea que el ancho de banda se triplica cada año; y, finalmente, la Ley de Metcalfe dice que el valor de una red es igual al cuadrado de los usuarios. Para una discusión sobre la Ley de Metcalfe, véase Krugman (2014).

⁶ En Francia, el costo de acceso a internet cayó de 75 dólares mensuales por una conexión *dial-up* en 1995 a 33 dólares mensuales por una conexión de banda ancha (mil veces más veloz).

⁷ Tasa de crecimiento anual compuesta.

⁸ Entre septiembre de 2001 y octubre de 2007, el costo por genoma cayó en un promedio anual del 21%; mientras que entre junio de 2008 y julio de 2014 lo hizo en un promedio anual del 44%. Véase <www.genome.gov/sequencingcosts/>.

procesamiento y almacenamiento (Barrantes, Jordán y Rojas, 2013), lo que redundará en un mayor desafío para crear valor a partir de los datos generados.

McKinsey (2011) identifica cuatro olas de adopción de tecnologías de la información (TI) en Estados Unidos con diferentes niveles de impacto en la productividad. También ha detectado que dicho impacto se da con cierto rezago, hecho que se explica por el tiempo que demanda incorporar las nuevas tecnologías en la organización del trabajo. Por lo tanto, es probable que las ganancias en productividad que se registran en un determinado período sean producto de la realización de las inversiones de períodos anteriores. De esta manera, puede que la totalidad de las bondades de Big Data no se hayan dejado ver todavía, aunque para los países en desarrollo implica el desafío adicional de cerrar la brecha tecnológica informacional y la capacitación de los recursos humanos. En el cuadro 1 se sintetizan los datos presentados por McKinsey (2011, pp. 24-25), donde se observa que los últimos dos períodos muestran las tasas de productividad por TI más elevadas, aun siendo los más cortos en duración—no superan los cinco años—, lo que refleja un progresivo aumento de la participación de TI en el crecimiento anual de la productividad desde los años setenta hasta el 2000, cuando muestra una desaceleración.

Cuadro 1 Etapas de adopción de tecnologías de la información en Estados Unidos
En porcentajes

Etapas		Período	Crecimiento anual de productividad	Contribución de TI	Participación de TI en productividad
Primera	Computadora central	1959-1973	2,82%	0,30%	11,00%
Segunda	Minicomputadoras y PC	1973-1995	1,49%	0,65%	44,00%
Tercera	Internet y web	1995-2000	2,70%	1,59%	59,00%
Cuarta	Dispositivos móviles y web 2.0	2000-2006	2,50%	0,96%	38,00%

Fuente: Elaboración propia con base en McKinsey [2011, pp. 24-25].

Las ganancias en productividad que puedan lograrse con Big Data dependen de las inversiones en TI—capacidad de almacenamiento y procesamiento—y de la innovación en la organización del trabajo que incorpore las nuevas herramientas. Para Barrantes, Jordán y Rojas (2013), Big Data es una de las tres trayectorias⁹ en las que se desarrolla la explosión en la generación de datos que protagonizamos por el aumento del uso de la tecnología en la vida diaria, cuyo aprovechamiento descansa y depende de la instalación de infraestructura de redes de alta velocidad.

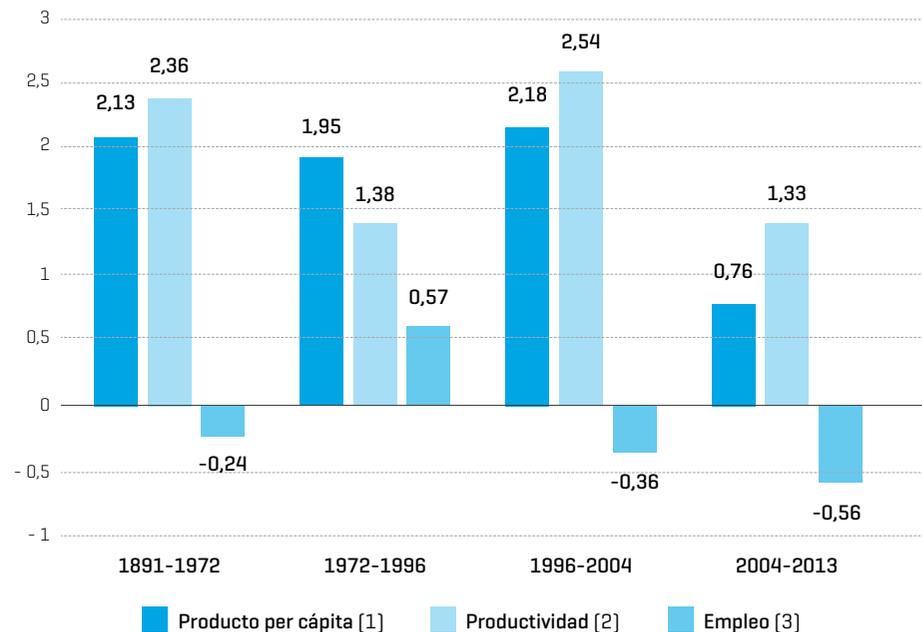
En el sentido de McKinsey (2011), UN Global Pulse, una división de Naciones Unidas para impulsar innovaciones en la velocidad de recolección y análisis de datos, vigente desde 2009, sugiere que nos encontramos frente a una revolución industrial de datos, precedida desde el comienzo del nuevo siglo por las innovaciones tecnológicas y los dispositivos digitales. Este período se caracteriza por un incremento exponencial en la cantidad y diversidad de datos digitales disponibles en tiempo real, producto de un mayor uso de equipos tecnológicos de mayor capacidad en la vida diaria, lo que permite alcanzar un mayor conocimiento del comportamiento humano (UN Global Pulse, 2012). Según las estimaciones que utilizan, la cantidad de datos digitales disponibles a nivel mundial pasaron de 150 exabytes en 2005 a 1.200 en 2010, esto es, un crecimiento de ocho veces, y se proyecta un crecimiento del 40% por año, es decir, cuarenta veces el crecimiento de la población mundial. Por lo tanto, no solo se estaría duplicando la cantidad de información digital acumulada cada veinte meses, reproduciendo una dinámica similar a la de la Ley de Moore, sino que también dichos datos se estarían haciendo cada vez más jóvenes en términos relativos, al haber una mayor disponibilidad en tiempo real. Asimismo, desde una perspectiva crítica sobre la naturaleza de las crisis internacionales, el documento sugiere que Big Data es una oportunidad, particularmente para los países en desarrollo, de evitar o moderar el impacto de los *shocks* financieros o climáticos y reducir el impacto de la volatilidad de los mercados, a partir de la disponibilidad de datos en tiempo real que habiliten a dar respuestas con mayor anticipación.

Sin embargo, al analizar el crecimiento económico y la productividad de Estados Unidos entre 1891 y 2013 (gráfico 1), Gordon (2014) observa que, luego de alcanzar un máximo en el período 1996-2000, la caída en el crecimiento del producto, la productividad y el empleo para el período 2004-2013 contradice la predicción de los tecno-optimistas de una aceleración

⁹ Las otras dos son la difusión de la banda ancha y la computación en la nube. Dado el alcance del presente documento, nos ocuparemos solo de Big Data, sin por ello considerar a estos dos de menor importancia.

del crecimiento económico. Durante los últimos diez años, el descenso de la productividad de la economía norteamericana no muestra ningún impulso provocado por la explosión de Big Data o la invención de *smartphones* o *tablets*. Para el autor, no hay sorpresas en estos resultados, ya que si los costos marginales se igualan a los beneficios marginales, en la medida en que el costo de los datos digitales tiende a cero, también lo harán los beneficios: ha dejado de existir la relación entre la Ley de Moore –o el crecimiento de los datos digitales–, por un lado, y el crecimiento del producto per cápita y la productividad, por el otro. Big Data es, para el autor, un juego de suma cero, ya que mayormente los datos son utilizados por grandes corporaciones para mercadotecnia y conductas predatorias por la participación en el mercado.

Gráfico 1 Producto per cápita, productividad y empleo en Estados Unidos [varios períodos]
En porcentajes

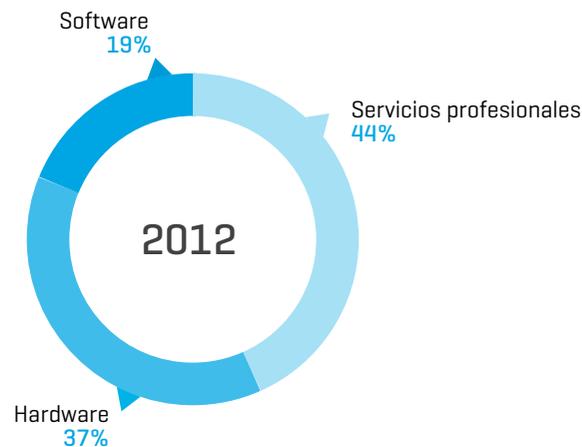


Nota: [1] PBI per cápita, [2] productividad por hora hombre medida en términos de la relación entre el PBI y las horas trabajadas y [3] empleo capturado a partir de la relación entre horas trabajadas y población.

Fuente: Gordon [2014].

En cuanto al mercado de Big Data, Kelly (2014) sostiene que en 2012 las ventas de hardware, software y servicios profesionales vinculados a Big Data crecieron un 59% respecto de 2011, y ascendieron así a 11.590 millones de dólares de acuerdo con un relevamiento realizado sobre más de sesenta grandes empresas. Para 2013 se esperaba un crecimiento del negocio en un 61% anual, que alcanzaría los 18.100 millones de dólares. En el gráfico 2 se señala la categoría "servicios profesionales" como aquella que mayores ingresos concentró. En este estudio, se observa que IBM registró ingresos por 1.250 millones de dólares, de los cuales el 50% se generó por servicios profesionales; mientras que HP, la segunda firma en importancia, generó 660 millones de dólares, cuyo 38% fue por servicios profesionales.¹⁰

Gráfico 2 Segmentación del mercado de Big Data (2012)



Fuente: Kelly (2014).

Si bien el mercado está dominado por grandes firmas como IBM, HP, Teradata y Dell,¹¹ es probable que las innovaciones sean impulsadas por pequeñas empresas dedicadas exclusivamente a Big Data, las cuales se vuelven objetivos de adquisición de las grandes compañías. Según la Comisión Europea (2013), esto obligará a las empresas a elegir entre los siguientes modelos de negocio: concesión de licencias del software que desarrollen, oferta de servicios profesionales, venta de sus productos y adquisición de la empresa por parte de otra más grande.

¹⁰ Véase el listado completo de firmas en Kelly (2014).

¹¹ De acuerdo con Kelly (2014), las firmas IBM, HP, Teradata y Dell concentran el 24% de los ingresos generados por actividades vinculadas a Big Data.

Consideraciones metodológicas y epistemológicas a partir de Big Data

Una discusión que se plantea desde la teoría sociológica, y que abarca tanto a las TIC en general y a Big Data en particular, tiene que ver con el rol que estas desarrollan dentro del orden social, más allá de la reflexión particular que se pueda hacer sobre ellas. Esto significa considerar a las TIC no como subproducto de las relaciones sociales o como factores tecnológicos, sino tomar en cuenta también su carácter sociológico y sus funciones comunicativas. Para ello, Forte *et al.* (2012) identifican tres obstáculos epistemológicos para el análisis de las TIC, que creemos es extensivo para la particularidad de Big Data:

Tecnologicismo. Reducción a sus características tecnológicas.

Sociologicismo. TIC como variable dependiente de la estructura social.

Sobrecarga diacrónica. Se imputan pretensiones diacrónicas a las TIC y se espera que devuelvan criterios para la distinción de etapas dentro de la evolución societal reciente, sin ponderarlas dentro de una teoría de la sociedad.

Como veremos más adelante en cuanto a la definición más aceptada de Big Data, estos obstáculos epistemológicos están presentes para un abordaje sociológico de Big Data. Para los autores antes mencionados, se trata de observar una estructura comunicativa históricamente novedosa ligada a tres procesos: digitalización –o informatización de contenidos–, convergencia tecnológica e internet. De esta manera, la digitalización supone la convergencia de distintos soportes técnicos que a su vez implican el robustecimiento y mundialización de la conectividad de internet, lo cual ofrece infraestructura y propagación dentro de la red. La digitalización se constituye en un medio de comunicación y en un umbral histórico a partir de su masificación, al transformarse en unidades trasladables y transmisibles de información mediante técnicas de binarización. La convergencia tecnológica se da dentro de esta digitalización y se reducen los grados de libertad del desarrollo de técnicas, lo que tiende a un acoplamiento tecnológico que forma una simplificación funcional o, en otros términos, una dependencia en la trayectoria de este desarrollo (*path dependence*). Incluso esta convergencia puede ser vista

como un fenómeno comunicativo porque se desarrolla en un medio de comunicación digital, pero al mismo tiempo condiciona la comunicación por la simplificación funcional, lo cual genera una dependencia mutua y el aumento del riesgo de disrupción digital/comunicacional, generalmente mencionado como brecha digital. Finalmente, internet ofrece la infraestructura telecomunicativa y su propagación a nivel mundial. Es por ello que la emergencia de un medio de comunicación digital transforma la comunicación y las estructuras de propagación de la comunicación.

Otra discusión vigente y de orden epistemológico tiene que ver con los cambios en el paradigma metodológico científico sobre los límites y la relación entre teoría y contraste empírico. En este sentido, Manovich (2012) se plantea cómo Big Data puede cambiar el método de investigación en las ciencias sociales y las humanidades, ya que en la medida en que el mundo se vuelve más y más digital, nuevas técnicas serán necesarias para investigar, analizar y entender esta cantidad de datos. Una primera manifestación de estos cambios es, para el autor, la división entre los métodos cuantitativos, comúnmente vinculados a las ciencias sociales, y los métodos hermenéuticos generalmente asociados a las humanidades: ya no sería necesario elegir entre el tamaño de la muestra y la profundidad de los datos. Asimismo, el autor advierte que para poder darle un uso provechoso a los datos masivos, primero hay que tener claro cuáles son nuestras limitaciones y, por lo tanto, cuáles deben ser las cualidades que los analistas de Big Data deben tener. Reivindica de esta manera el rol del analista –particularmente de los humanistas– frente a una completa automatización por métodos computacionales a la hora de la limpieza e interpretación de los datos y la detección de patrones. Si bien mantiene un enfoque optimista sobre el potencial de esta herramienta, considera que todavía se encuentra en una etapa embrionaria, donde las empresas que recolectan Big Data utilizan la información para detectar patrones de comportamiento y ofrecer productos, personalizar publicidad, etc. Estas empresas no comparten estos datos, cuanto mucho ofrecen información sintetizada.

Por su parte, danah boyd y Kate Crawford (2012) consideran a Big Data como un fenómeno cultural, tecnológico y académico, que descansa en la interacción entre la tecnología

—maximizando la capacidad computacional y la precisión de los algoritmos—, el análisis para identificar patrones y facilitar la toma de decisiones y la mitología o creencia que grandes cantidades de datos ofrecen una forma de conocimiento más elevada y una mejor aproximación a la verdad. Como todo nuevo fenómeno sociotecnológico, Big Data despierta posiciones utópicas y catastróficas sobre la nueva etapa. La visión optimista sugiere que Big Data permitirá resolver graves problemas sociales en las áreas de salud, terrorismo o inseguridad, cambio climático, etc.; mientras que la visión pesimista alerta sobre el advenimiento de una era "gran hermano", con violación a la privacidad de las personas, menos libertades civiles y mayor control corporativo. Sin embargo, las autoras destacan que las posiciones extremas oscurecen la discusión sobre los cambios vertiginosos que este nuevo fenómeno protagoniza en un contexto de incertidumbre: las decisiones que se tomen en el surgimiento de esta etapa condicionarán su desarrollo futuro.

Al considerar la creciente capacidad de recolección de datos y de algoritmos capaces de determinar patrones de comportamiento humano, es necesario discutir qué sistemas están motorizando estas prácticas y cuáles los regulan. En el trabajo de Boyd y Crawford, el fenómeno Big Data adopta una jerarquía semejante al fordismo: este nuevo fenómeno emerge como un sistema de conocimiento que está cambiando incluso los mismos objetos de conocimiento,¹² la epistemología y la ética, y plantea una transformación de la teoría y el aprendizaje a partir de nuevas herramientas. En la misma línea de Manovich (2012), las autoras plantean que la disponibilidad de grandes cantidades de datos renueva la discusión sobre objetividad y subjetividad del método científico, particularmente en las ciencias sociales y las humanidades. La cantidad de datos por sí sola no otorga objetividad absoluta, ya que estos se encuentran sujetos a interpretación, del mismo modo en que el objeto observado y las decisiones de investigación se basan en la subjetividad del investigador. Big Data no se explica por sí solo y los métodos de interpretación están sujetos a debates filosóficos.¹³ Esto es así aun sin considerar los errores que puedan tener los datos masivos, la dificultad de interpretación que se genera por fuentes desconocidas, la (poca) representatividad del conjunto de datos elegido o su descontextualización, entre otras cuestiones. Una mayor cantidad de datos

¹² Chen y Zhang (2014) plantean que el fenómeno Big Data, que da lugar a la ciencia de datos, se presenta como un cuarto paradigma científico, debido a los cambios que las crecientes aplicaciones intensivas en datos están provocando en la ciencia. Los otros tres paradigmas serían la ciencia empírica (explicación de la naturaleza basada en experiencia empírica), la ciencia teórica (Leyes de Newton y Kepler) y la ciencia computacional (simulación científica).

¹³ En términos más generales, la realidad no se explica por sí sola y, por lo tanto, un empirismo radical no es sostenible, ya que forma parte de un sistema. De esta manera, un dato ya es una construcción sobre aquello que es observable; es decir, no existe un dato crudo absolutamente puro y objetivo (Forte, 2015).

no necesariamente significa mayor calidad; es decir, los datos no quedan exentos de la necesidad de mecanismos de validación de los resultados.

El aspecto ético abordado en el documento tiene que ver con la disponibilidad y el uso de los datos de las personas. Que la información esté disponible o parezca pública no implica que no deban resguardarse los derechos de las personas. Estar en público –por ejemplo, estar en un parque– no es lo mismo que la exposición pública –por ejemplo, reclamar activamente atención–. Finalmente, las autoras también destacan el problema de la desigualdad en cuanto al limitado acceso a Big Data y la brecha digital; asimismo, plantean la discusión sobre la apropiación de los datos por parte de las empresas y la problemática de las dimensiones pública y privada.

Por su parte, David Bollier (2010) aborda la discusión metodológica sobre cómo deberían ser elaboradas las teorías en la era Big Data y la importancia de la interpretación de las correlaciones y predicciones que se realizan a partir del procesamiento de datos masivos. Hace cinco años, la herramienta *Google Flu Trends* sorprendió con la aparente posibilidad de predecir epidemias en forma rápida, precisa, económica y sin teoría. De hecho, el estudio ni siquiera planteaba una hipótesis (Wired, 2008). Sin embargo, cuatro años más tarde, las predicciones del modelo libre de teoría de Google Trends mostró grandes fallas de predicción. El problema fue que Google no analizó cuáles eran los factores que vinculaban los términos de búsquedas con la propagación de la gripe; se ocuparon de correlaciones sin preocuparse por la causalidad. Grandes cantidades de datos no solucionan por sí mismos los problemas estadísticos que llevan años de investigación (Harford, 2014; Lazer *et al.*, 2014). A esto debemos agregar la representatividad de los datos, como lo demuestra el hecho de que en África apenas más del 16% de la población tiene acceso a internet, o el fracaso que resultó aplicar *Google Flu Trends* en Bolivia, donde la población, a diferencia de lo que ocurre en Estados Unidos, confía más en la consulta al médico para el diagnóstico de la gripe antes que en la búsqueda en internet (SciDev. Net, 2014). En este sentido, Martin Hilbert (2013b) advierte las limitaciones que puede tener Big Data para el desarrollo socioeconómico y rechaza las posturas apologéticas que plantean el fin de la teoría. El análisis de Big Data aporta información sobre la estructura

y la dinámica de hechos pasados, por lo que su capacidad de predicción queda sujeta a que no haya cambios significativos. Sin embargo, el objetivo de las políticas de desarrollo consiste en crear un nuevo escenario, diferente al pasado, cambiando la lógica de funcionamiento del sistema socioeconómico y cultural. Dada la complejidad que esto reviste, por un lado vale cuestionar cuánto puede decir el pasado sobre el futuro y, por el otro, se debe implementar modelos basados en la teoría, cuyos grados de libertad permitan realizar los ajustes necesarios para la predicción de comportamientos. Hilbert propone el uso de Big Data en tiempo real en modelos de simulación computacional –como los modelos basados en agentes–, cuya capacidad de adaptación permitan ir de la teoría general a los casos concretos.

Desde la perspectiva del desarrollo económico-social, Hilbert (2013a) adopta un esquema analítico tridimensional entre tecnología, cambio social y políticas estratégicas, que procede de la literatura del ICT4D (Tecnologías de Información y Comunicación para el Desarrollo) y que está basado en la noción schumpeteriana de evolución social a través de la innovación tecnológica. Este autor considera que Big Data aplicado a la toma de decisiones va a tener un efecto positivo sobre la eficiencia y la productividad, similar al que mostraron las TIC en las décadas recientes (Peres y Hilbert, 2010), abriéndose los mismos problemas para los países en desarrollo.¹⁴ Por lo tanto, aun cuando esto pueda redundar en una mejora en la toma de decisiones de áreas clave como la salud, el empleo, la productividad, la seguridad y el manejo de los recursos naturales, Hilbert advierte que cuestiones tales como la privacidad y los aspectos técnicos/tecnológicos se encuentran agravados en los países en desarrollo por otros factores, como el atraso tecnológico, la falta de infraestructura y la carencia de personas capacitadas. Este nuevo paradigma se desarrollaría a través de un proceso de difusión lento y desigual, comprometido por la falta de infraestructura, de capital humano y la ausencia de disponibilidad de recursos económicos e institucionales en los países en desarrollo. Esto abre una nueva brecha digital basada en el análisis de datos para la toma de decisiones inteligentes. En consecuencia, la mejora en el desarrollo económico y social no se da automáticamente, sino que requiere elaborar estrategias y políticas públicas específicas. En esta misma dirección, UN Global Pulse (2012, p. 6) considera que Big

¹⁴ En los países en desarrollo el problema con las TIC es que la demanda es poco sofisticada y por lo tanto se limita la emergencia de una oferta de mayor complejidad. Buena parte de las soluciones suelen ser de tipo incorporado que se generan en países desarrollados.

Data aplicado al desarrollo significa convertir datos imperfectos, complejos y desestructurados en información procesable, aunque esto no se da automáticamente.

De acuerdo con estos enfoques, más allá de la escala de Big Data, la clave del nuevo paradigma es el análisis sistemático de datos masivos, previo a y para una mejor toma de decisiones. Este proceso podría ser entendido como un paso natural en la evolución de la era de la información o de las sociedades de la información —en términos de Bell, Castells y otros autores— hacia las sociedades del conocimiento. Basadas en la infraestructura digital que permite un vasto incremento de la información, el desafío actual consistiría en convertir la información digital en conocimiento para la toma de decisiones inteligente (Hilbert, 2013a).¹⁵

Por lo tanto, si se acepta la posición que plantea estar en presencia de un cambio paradigmático en la forma de generar la información y transformarla en conocimiento, no debemos dejar de mencionar algunos aspectos estructurales de la creación y apropiación de valor a partir de la información. Por un lado, es necesario destacar que la información, mientras más se sociabiliza, más se valoriza; esto es, mientras haya más personas trabajando con un determinado conjunto de datos, mejores y más resultados se pueden obtener. Sin embargo, esta proposición entra en contradicción con la lógica de acumulación vigente, caracterizada por una frágil inclinación a la libre difusión de la información y el conocimiento. Los datos, la información y el conocimiento son activos sociales que no se agotan con el uso y, salvo por el soporte físico que necesiten para estar disponibles, tampoco son exclusivos, por lo que lo único que puede imponer una limitación en la difusión de estos son los medios sobre los que se diseminan —el soporte tecnológico—. Quien accede a controlar los medios, puede controlar los contenidos. En el caso particular de Big Data, es importante señalar que quien genera el dato no necesariamente es quien lo almacena: el 70% del total de los datos mundiales está almacenado en Norteamérica y Europa, lo que se expresa como brecha digital o informacional.¹⁶ Lev Manovich (2012) sugiere que este nuevo fenómeno crea una nueva división de clases o data-clases en la sociedad: aquellos quienes generan los datos, es decir, la mayoría de la población;

¹⁵ Los autores que menciona Hilbert (2013a) forman parte de un debate acerca de la caracterización y denominación de las distintas eras o sociedades, que no será abordado en el presente trabajo. Sin embargo, vale destacar que, si bien se menciona el paso hacia un nuevo tipo de sociedad, no consideramos que dicho tránsito sea automático, sino que es necesario un proceso *path dependence*.

¹⁶ Para una discusión vigente en América Latina, véase *Ámbito Financiero* (2014).

aquellos que tienen los medios para recolectarla, que son muy pocos; y aquellos que tienen la capacidad para analizarla, que representan un sector de la sociedad aun menor. Existen, por lo tanto, jerarquías y relaciones de poder, debido a que quienes poseen los medios para captar los datos tienen el poder para apropiarse de la renta generada. boyd y Crawford (2012) aceptan esta división en data-clases y agregan la distinción entre "Big Data ricos" y "Big Data pobres" al interior de los sectores que acceden de una u otra forma a los datos masivos—por ejemplo, las universidades.

En principio, una alternativa es hacer públicos los datos; sin embargo, esto abre al menos otros dos planteos: cómo hacer públicos los datos que se relevan de forma privada y, segundo, quién recopila los datos. La recopilación y limpieza de los datos es parte del proceso de análisis, donde la objetividad del dato en bruto ya se pierde y queda condicionada por la forma en que se recopila y selecciona, por lo tanto, condicionada a los intereses de quien realiza esta tarea, lo cual no necesariamente está ajustado a la política de desarrollo adoptada por el país o la región. Al respecto, UN Global Pulse (2012) plantea entre los desafíos para el desarrollo dimensiones tales como: implicancias conceptuales, legales y tecnológicas de la privacidad; el acceso y la disponibilidad—la reticencia de las empresas privadas y otras instituciones para compartir datos sobre clientes, usuarios, operaciones, etc.—; y el análisis acerca de lo que está sugiriendo el dato—debemos agregar el desafío de obtener el dato correcto, darle sentido y detectar las anomalías. Los desafíos aquí planteados están sujetos a todo tipo de subjetividad. Un caso importante y estratégico para los países en desarrollo es el de los recursos naturales, donde la tecnología para la explotación suele estar en manos privadas—nacional o extranjera—y por lo tanto el uso monopólico de la información puede dar lugar a la apropiación privada de rentas informacionales.

Definición de Big Data y aspectos técnicos

Una primera aproximación al tema Big Data nos lleva a preguntarnos qué es Big Data. Dada la novedad, el tema todavía no se halla suficientemente delimitado, aunque

podemos encontrar varios puntos en común entre las definiciones planteadas por la comunidad académica, los técnicos y los profesionales que se encuentran indagando sobre el tema.

Una definición técnica que se desprende del nombre alude a la escala o volumen del conjunto de datos. McKinsey (2011, p. 1) ofrece una definición intencionalmente subjetiva y dinámica, a partir de las limitaciones tecnológicas de un determinado período: Big Data refiere a un conjunto de datos cuyo tamaño está más allá de la capacidad que tienen los software de base de datos para capturar, almacenar, administrar y analizar. De esta manera, a medida que la frontera tecnológica se va modificando, el tamaño del conjunto de datos que califica como Big Data también irá cambiando. Hablamos de Big Data cuando el tamaño se vuelve parte del problema. Si bien el volumen es un punto en común, otras definiciones incorporan más dimensiones al análisis. Algunas compañías que investigan sobre la materia¹⁷ suelen caracterizar los desafíos que Big Data genera no solo a partir del volumen, sino también de la variedad y la velocidad (3Vs), incluso dándole mayor importancia a estas últimas. Con variedad se suele hacer referencia a la heterogeneidad de la representación e interpretación semántica de los datos, es decir que no están directamente listos para ser integrados a una aplicación;¹⁸ mientras que con velocidad—o *timeliness*, según el término en inglés—se hace referencia tanto a la frecuencia de disponibilidad de los datos como al tiempo en el que hay que dar respuesta. La importancia reside en la velocidad del *feedback*, utilizando los datos para la toma de decisiones. En algunas ocasiones también se incluye una cuarta V que remite a valor, variabilidad o virtual (Armah, 2013; OECD, 2013; Chen y Zhang, 2014). A pesar de la importancia de estas características, según la publicación UN Global Pulse (2012) restaría incluir privacidad de las personas¹⁹ y facilidad de uso (*usability*).

Es decir, Big Data son datos que exceden la capacidad de procesamiento de los sistemas de base de datos convencionales, caracterizados por un gran volumen, por la gran velocidad con la que se transmiten y por no ajustarse a las restricciones de las bases de datos convencionales. La definición de Big Data aceptada al momento se funda en las características y limitaciones que este nuevo paradigma tecnológico informacional impone

¹⁷ Por ejemplo, Gartner (2011) y UN Global Pulse (2012).

¹⁸ Algunas estimaciones consideran que los datos no estructurados que alcanzan entre el 80% y el 85% de los datos totales están mayormente subaprovechados (OECD, 2013).

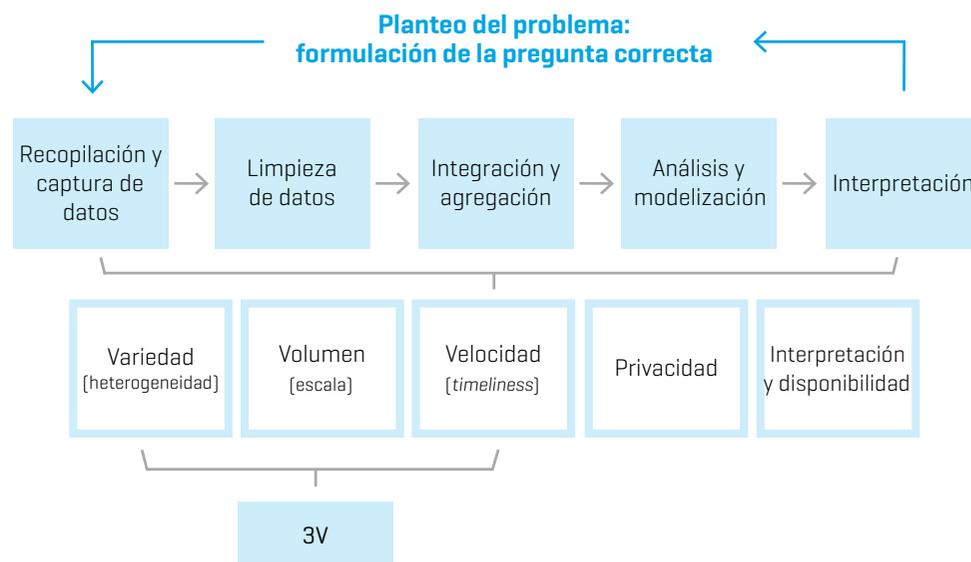
¹⁹ Según Cumbley y Church (2013), el análisis de Big Data consta de cuatro etapas: recolección, combinación, análisis y uso. Para analizar el riesgo de violación de la privacidad de las personas debemos comenzar por la etapa de recolección de datos, que abarca: las comunicaciones electrónicas—por ejemplo, los correos electrónicos—, el internet *tracking*—como las *cookies*—, la tecnología para identificación de radiofrecuencia—por ejemplo, la Octopus Card en Hong Kong, la tarjeta Oyster en Londres, la SUBE (Sistema Único de Boleto Electrónico) en la Argentina—, la geolocalización a través de *smartphones* y *tracking* vehicular—como el GPS—, la videovigilancia—por ejemplo, las cámaras en espacios públicos y privados—, la información financiera—como el consumo con tarjeta de crédito, el pago electrónico, etc.—y la conservación de registros electrónicos—por ejemplo, la nube digital para almacenar información.

sobre las capacidades técnicas y cognitivas actuales. Crear valor a partir de Big Data exige cambios en la forma de procesar los datos, de pensar y de plantear los problemas.

Análisis de Big Data. Un esquema conceptual

Siguiendo a Hilbert (2013a), lo crucial de Big Data no es el gran aumento de los datos en sí mismo, sino su análisis para la toma de decisiones inteligente, por lo que prefiere la expresión "análisis de Big Data" a la de "Big Data". En este sentido, boyd y Crawford (2012) consideran que la novedad de Big Data, más que el volumen, se encuentra en las capacidades de búsqueda y agregación de grandes cantidades de datos relacionados (*cross-reference data sets*). Por otro lado, sin diferir en lo sustancial con lo anterior y ofreciendo un esquema analítico más claro a los fines conceptuales,²⁰ UN Global Pulse (2012) sugiere que el

Figura 2 Secuencia de análisis de Big Data [2012]



²⁰ Otras consideraciones sobre el esquema de Big Data incorpora el uso de las herramientas vigentes Map Reduce, SQL, NoSQL, entre otras. Véase Tekiner y Keane (2013).

Arriba: las principales etapas en el análisis de Big Data. **Abajo:** las características de Big Data que hacen de esta tarea un desafío.

Fuente: Elaboración propia con base en UN Global Pulse [2012].

análisis de Big Data involucra múltiples fases que plantean un desafío en sí mismo y proponen un esquema lineal para caracterizar todo el sistema, al que agregamos un proceso de retroalimentación (figura 2).

Si bien la etapa de análisis y modelización (*analysis/modelling*) suele ser la que mayor atención atrae según los autores, no se obtienen demasiados resultados si se descuidan el resto de las etapas y el contexto en el cual se realiza el análisis de Big Data. La fase de filtrado y limpieza de los datos puede demandar hasta el 80% del trabajo de análisis.²¹ Si aceptamos esta estimación y tenemos en cuenta que una vez que el dato es manipulado se pierde su objetividad intrínseca, prácticamente no hay objetividad posible. Por lo tanto, quien tiene la propiedad de los datos—o los medios para apropiarse de ellos—y quien realiza el análisis es determinante en la construcción de sentido.

Esto agrega otro desafío en el análisis de Big Data: no solo plantea un problema técnico de hardware y software, sino que también requiere técnicos y profesionales capacitados para llevar a cabo esta tarea. De acuerdo con Peter Sondergaard, vicepresidente de Gartner, empresa dedicada a la investigación de TI, en 2013 se pronosticaba una demanda de 2 millones de científicos de datos²² en los próximos tres años solo para Estados Unidos—y de 4 millones para todo el mundo—, de los cuales solo un tercio serían cubiertos. Incluso, sugería que cada puesto relacionado con Big Data en Estados Unidos generaría tres empleos por fuera de TI, ofreciendo 6 millones de nuevos puestos de trabajo en tres años (Fisher, 2013).

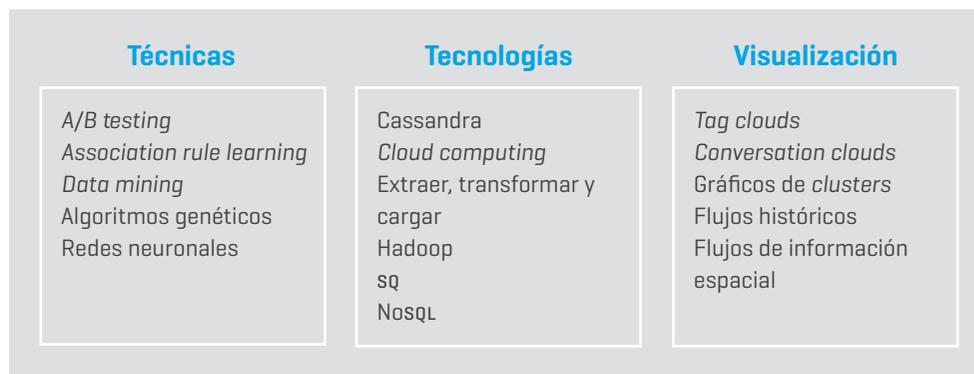
Existe una gran variedad de técnicas y tecnologías para agregar, manipular, analizar y visualizar Big Data, desarrolladas a partir de diferentes campos como la estadística, las ciencias de la computación, la matemática aplicada y la economía, a partir de un enfoque flexible y multidisciplinario.²³ La Comisión Europea (2013) elaboró la lista de técnicas, tecnologías y herramientas (figura 3):

²¹ Véase <<http://strata.oreilly.com/2012/01/what-is-big-data.html>>.

²² La ciencia de datos aparece como una nueva profesión vinculada a las ciencias matemáticas, estadísticas y computacionales, con perfil analítico y capacidad de un abordaje multidisciplinario de los problemas.

²³ Para un detalle de las técnicas, tecnologías y herramientas de visualización de Big Data vigentes, véase McKinsey (2011, pp. 27-36). Asimismo, véase el Apéndice (p. 37 del presente documento).

Figura 3 Diferentes tipos de técnicas, tecnologías y herramientas de visualización de Big Data



Nota: Cada una de estas técnicas, tecnologías y herramientas se desarrolla en el Apéndice de este documento.

Fuente: Comisión Europea [2013].

Big Data: aplicaciones sociales y económicas. Oportunidades para el desarrollo

En este apartado se relevan en forma aún preliminar los sectores que, de acuerdo con la literatura, muestran experiencias en el uso de Big Data.

Salud

McKinsey (2011 y 2013) ha estudiado el potencial impacto de Big Data en el incremento de la productividad del sector salud en Estados Unidos. Este sector fue elegido por la participación en el producto bruto interno (PBI) de Estados Unidos —17% en el 2010—, por la necesidad del sector de aumentar la productividad. En la última década, el gasto en salud crecía el 5% anual y en 2007 el gasto per cápita ajustado por poder adquisitivo era 30% superior al promedio de los países de la OCDE, sin mostrar mejores resultados en los

cuidados de la salud; y por la relevancia de las partes involucradas, como la industria farmacéutica y de productos médicos, los proveedores y los pacientes.

Considerando las inversiones necesarias en tecnología, personal capacitado, cuidado de la privacidad y con los incentivos adecuados, el estudio estimaba que las ganancias potenciales con Big Data en diez años ascenderían a 300 mil millones de dólares anuales, cuyos dos tercios se explican por reducciones en los gastos de atención de la salud. Esto representa el 8% del nivel de gasto estimado en salud para 2010. Las estimaciones conservadoras de productividad sugieren una mejora del 0,7% anual a partir de la reducción de costos por tratamientos (drogas) poco efectivos, reducción de efectos adversos provocados por tratamientos, reducción de errores médicos y criterios sesgados hacia intervenciones poco efectivas, detección de patrones en las patologías, monitoreo remoto de los pacientes y sistemas personalizados. Los principales desafíos son: la digitalización de gran parte de la documentación (por ejemplo, las historias clínicas) y la manipulación de imágenes, la privacidad, la compartimentación de la información entre las empresas y organismos públicos, y el personal capacitado.

Por otro lado, el trabajo menciona algunas aplicaciones de Big Data en la salud, en Estados Unidos, el Reino Unido e Italia, que mostraron buenos resultados, principalmente vinculados al efecto adverso y efectividad de medicamentos y tratamientos.

El uso de Big Data puede mejorar el monitoreo de la salud pública y la capacidad de respuesta a partir de una base de datos de pacientes y tratamientos que permita una coordinada y temprana detección de enfermedades infecciosas, como así también reducir los tiempos de investigación. La revista *Wired* (2010) muestra los resultados de una investigación sobre el mal de Parkinson alcanzados por el modelo tradicional de investigación del National Institute of Health y el modelo empleado por la organización Parkinson's Genetic Initiative. El primer modelo sigue el esquema tradicional de planteo de hipótesis, estudios, recolección de resultados, análisis, elaboración del documento, presentación del documento en *The New England Journal of Medicine* y publicación. El otro modelo sigue una secuencia diferente: preparación de la herramienta, en este caso, una encuesta;

reclutamiento de pacientes, recolección de resultados, análisis y presentación de resultados en el encuentro de la Royal Society of Medicine en Londres, a la espera de publicar un documento más tarde. Lo interesante es que ambos estudios llegaron prácticamente a los mismos resultados, aunque al tradicional le llevó seis años y al segundo solo ocho meses.

Hilbert (2013a) cita un estudio de caso en Estados Unidos sobre atención médica de lesiones de cadera y de antebrazo, cuyas conclusiones permiten identificar cuatro tipos de variaciones en el tratamiento y afección de determinadas patologías: patrón geográfico-ambiental, errores médicos, criterios sesgados para determinar la necesidad de cirugía y sobreutilización y sobreoferta de intervenciones y gastos en algunas regiones por abundancia.

Harding y Lovenheim (2013) estudian el rol de los precios en el patrón de consumo y nutrición en Estados Unidos entre 2000 y 2007, y combinan técnicas de demanda estructural con el enfoque de Big Data.²⁴ Esto les permite caracterizar las decisiones de consumo e investigar el efecto de diferentes impuestos a los alimentos para mejorar el impacto nutricional del consumo alimenticio y prevenir problemas de obesidad. La conclusión es que el impacto de una política impositiva sobre los componentes nutricionales es más efectivo sobre el consumo alimenticio saludable que la aplicación de impuestos sobre los productos.

Administración pública

McKinsey (2011) estudia cómo incrementar la productividad del sector público en la Unión Europea, que representa el 50% del PIB y cuyo principal factor a tener en cuenta es el envejecimiento de la sociedad que imprime presión sobre el déficit fiscal. Si bien el sector público genera menos datos—medidos en bytes—respecto del sector de la salud, la ventaja es que el 90% de esta información está digitalizada. El estudio se centralizó en las agencias de recaudación de impuestos y de trabajo. De acuerdo con los cálculos, el sector público en 23 de los estados más grandes de Europa puede crear potencialmente entre 150 y 300 mil millones de euros en los próximos diez años, lo que representaría—de mantenerse la tendencia— un aumento de la productividad del 0,5% anual. Las tres dimensiones

²⁴ El enfoque de Big Data que utilizan estos autores se basa en la vinculación conjunta de varias bases de datos grandes para proporcionar una visión global de consumo, precios, características nutricionales de los alimentos comprados y el entorno en el que se realizó la compra.

identificadas para este aumento son: la eficiencia operativa, la reducción de errores y fraudes en la administración de beneficios, y la reducción de la brecha fiscal—recaudación potencial menos recaudación real— a partir de un incremento de la recaudación de impuestos. Un factor que puede ayudar es incrementar la transparencia, mientras que uno de los principales desafíos metodológicos es medir la productividad en este sector.

Por otro lado, Einav y Levin (2013) destacan que el sector público es una gran fuente de datos en sectores tales como la seguridad social, el sistema impositivo, los programas sociales y la educación. Estas fuentes están subutilizadas, aunque hay algunos casos positivos en países europeos donde se releva información demográfica individual, situación de empleo y salud de la totalidad de la población. Cita también un trabajo de inequidad en la distribución del ingreso de Piketty y Saez (2003) y la situación del percentil más rico de Estados Unidos entre 1913 y 1998, a partir de los datos del Internal Revenue Service, como potencial aplicación de esta herramienta.

Por su parte, Hilbert (2013a) menciona aplicaciones con referencias geográficas (GPS) para la administración del tránsito, útiles para el transporte público, policía, bomberos, planificación urbana, etc. En Estocolmo, taxis y camiones provistos con GPS envían información cada 30 y 60 segundos, y ofrecen un panorama del tránsito en tiempo real que, sumado a los datos históricos y climáticos, permite elaborar proyecciones sobre las condiciones del tránsito (Biem *et al.*, 2010). También este tipo de relevamiento sirve para hacer un mapeo del delito como ya fue aplicado en Chicago. En este documento se halla también una discusión sobre el potencial aprovechamiento de los datos por el sector público y sobre el resguardo de la privacidad.

Recursos naturales, medio ambiente y energía

Hilbert (2013a) aborda la discusión de la difusión de la información y la competitividad a partir de los casos registrados en recursos naturales. La evidencia sugiere que una mayor disponibilidad de los datos, que es posible recolectar en las áreas del petróleo, el gas y la industria minera, aumenta la transparencia y no afecta la competitividad de la actividad.

En este trabajo, el autor también menciona aplicaciones al medio ambiente, prevención de desastres climáticos y alerta temprana (United Nations University, 2012), cuellos de botella en recursos naturales y aplicaciones concretas de seguimiento del clima y el comportamiento social para comercio minorista.

Otro sector para tener en cuenta respecto de los recursos naturales es el agro y la agricultura de precisión. Este caso ya está mostrando desarrollos en la Argentina de la mano del proyecto de la Fundación Sadosky con la Asociación Argentina de Consorcios Regionales de Experimentación Agrícola (AACREA). La agricultura de precisión permite realizar un mapeo del suelo al obtener datos sobre componentes químicos, humedad, etc. También ya existen aplicaciones de sensores y Big Data para la producción de alimentos y biotecnología (Carvalho *et al.*, 2014; Rubens, 2014).

Dada la dependencia de combustibles fósiles importados en Taiwán, Chou y Telaga (2014) realizan un estudio experimental de 17 semanas, donde desarrollan una metodología para predecir y detectar anomalías del consumo eléctrico en edificios de oficina en tiempo real. Los datos pueden ser recogidos por medidores inteligentes (*smart meters*) en tiempo real. La metodología consta de dos etapas: una de predicción que permite la planificación del consumo energético y otra de detección de anomalías que permite identificar patrones de comportamiento. Ambos contribuyen a la eficiencia en el consumo energético.

Industria manufacturera

McKinsey (2011) considera que si bien la industria manufacturera ha registrado altos niveles de productividad y eficiencia durante la década de 1990, todavía puede alcanzar mejoras a través de Big Data en la eficiencia de diseño y producción, la calidad del producto, la comercialización, promoción y distribución. Según el estudio, la consultora identificó siete niveles potenciales en los que el sector puede obtener mejoras gracias a Big Data a través de toda la cadena de valor y que pueden agruparse en cuatro grupos: inversión y desarrollo; reducción de la incertidumbre a través de proyecciones de

demanda y planeamiento de la cadena de suministro; eficiencia en la producción a partir de técnicas de simulación, sensores, etc.; y estrategias de comercialización y servicios posventa.

Transporte

Li *et al.* (2014) utilizan grandes volúmenes de datos históricos que pueden ser recolectados en tiempo real por sensores para aumentar la velocidad de la red ferroviaria y dar respuesta al aumento de la demanda del servicio. Combinando datos históricos con datos de fallas, de mantenimiento y climáticos, buscan elaborar modelos para predecir fallas en el servicio—por ejemplo, descarrilamientos— como forma de aumentar la velocidad de la red. Para ello utilizan técnicas de aprendizaje automático (*machine learning*) que permiten diseñar un sistema de alarma para responder ante la probabilidad de fallas, y así ganar en seguridad y mejorar los tiempos del servicio.

Otras de las aplicaciones más difundidas y con mayor potencial de Big Data es la optimización del transporte en sentido amplio, desde transporte público urbano hasta aerocomercial. De acuerdo con un relevamiento realizado por la Comisión Europea (2013), en Francia ya se ha implementado el uso de Big Data al proporcionar herramientas de análisis en tiempo real para minimizar los costos operativos asociados con el mantenimiento no programado de las aeronaves.

Otras alternativas que ofrece Big Data

En el UN Globe Pulse (2012) se mencionan otras aplicaciones de Big Data que enumeramos a continuación: astronomía (SDSS-III, 2008) y Big Data de imágenes; sector financiero y análisis de riesgo financiero sistémico a partir de compartir datos que las entidades recogen digitalmente; medio ambiente y difusión de sensores para realizar avances en cuestiones tales como los cursos de ríos y las emisiones de humo en incendio de reservas; bioinformática y nuevas tecnologías de secuencias (*Next Generation Sequencing*), que

permiten alcanzar más cantidad y calidad de datos genéticos; mejora de calidad y reducción de costos en el área salud; planeamiento urbano; transporte inteligente; seguridad y crimen organizado.

Nii Ayi Armah (2013) hace un relevamiento para el Banco de Canadá del uso potencial de grandes volúmenes de datos para estimar variables económicas tales como el PBI, la inflación y el desempleo, con una mayor frecuencia que la que permitirían los índices oficiales y tradicionales. Considerando que la población tiende a utilizar cada vez más medios de pago electrónicos y que hay una disponibilidad de datos digitales tempranos para el análisis, se podría estimar el nivel de actividad económica con mayor frecuencia. Entre los desafíos, destacan aspectos metodológicos, de acceso a los datos y de privacidad de las personas.

Choi y Varian (2011) hacen una revisión sobre aplicaciones de Google Trends para *now-casting*, esto es, predecir el presente, a diferencia del *forecasting* que intenta predecir el futuro. Entre las destacadas y más conocidas se encuentran la predicción de la tasa de desempleo a partir de las búsquedas en la web, la demanda de automóviles, las epidemias, los destinos turísticos y la inflación. También realizan un estudio propio para los casos de destinos de viajes, desempleo, confianza del consumidor y demanda de automóviles.

UN Global Pulse (2012) menciona aplicaciones de una metodología desarrollada por la International Food Policy Research Institute para detectar volatilidades de precios en los alimentos que pongan en riesgo la seguridad alimentaria de determinados países. También destaca un sistema aplicado a la detección de la amplitud de sismos y terremotos o epidemias, basado en mensajes de Twitter. Este organismo lleva adelante un proyecto con una firma privada de análisis de medios y redes sociales para estudiar el comportamiento de las poblaciones de Estados Unidos e Indonesia sobre determinados tópicos como vivienda, alimentos y petróleo, entre otros. Entre las virtudes que predominan del estudio del comportamiento humano, esta organización sugiere que se pueden avanzar en tres áreas: sistemas de alerta temprana, alertas en tiempo real y retroalimentación en tiempo real e impacto de las políticas públicas.

Conclusiones

Big Data surge como una nueva fase del paradigma intensivo en comunicación e información que emerge a partir de la revolución tecnológica iniciada en la década del setenta. En este sentido, si bien ofrece oportunidades para alcanzar tanto mejoras en competitividad y productividad como en opciones para la mejora en la calidad de vida, con potenciales beneficios para los países en desarrollo, se requieren importantes desarrollos en el área de TIC. En efecto, Big Data abre oportunidades, pero el análisis de datos masivos para la toma de decisiones inteligentes pone de manifiesto la necesidad de generar capacidades para superar la brecha digital.

Por lo tanto, la materialización de los potenciales beneficios para los países en desarrollo requiere la elaboración de políticas activas y específicas que tengan en cuenta la generación y apropiación de rentas informacionales, el manejo de la privacidad en cuanto a los datos personales, el desarrollo de capacidades para la creación de valor y la difusión de la información y el conocimiento para contribuir a la reducción de desigualdades socioeconómicas.

Si bien necesitaremos profundizar en los sectores de aplicación de Big Data en la Argentina, a continuación suministramos una lista preliminar con potenciales áreas de aplicación:

Sector agropecuario. Los avances tecnológicos y científicos del sector lo convierten en una potencial área de aplicación. De hecho, la Fundación Sadosky y AACREA están desarrollando actividades en conjunto. Sin embargo, no hay que descartar las brechas de productividad que existen con el resto de los sectores, quienes se apropian de la renta informacional y la sustentabilidad del modelo de negocio que se aplica al sector.

Bioteología. El potencial de Big Data en el área de la salud es alentador. La Argentina tiene desarrollos e investigación en secuenciación genómica, medicina personalizada, bioinformática, etc. El Instituto de Química Física de los Materiales, Medio Ambiente y

Energía, de la Facultad de Ciencias Exactas y Naturales de la Universidad de Buenos Aires, ha realizado estudios y aplicaciones al respecto. Los sectores de implementación pueden ser agropecuario –medicina veterinaria, genoma de especies–, salud –medicina personalizada–, seguridad –perfil genético– y biotecnología de combustibles.

Clima. Dada las herramientas satelitales del país y las capacidades desarrolladas por ARSAT, se podrían obtener datos para combinarlos con el sistema de transporte y logística, prevención de desastres naturales, planificación urbana, agricultura, etcétera.

Deporte. Existe actualmente el análisis de datos masivos aplicados al deporte, como fútbol, tenis, atletas en general y otras disciplinas. Con el impulso que ha recibido el deporte argentino en los últimos años, esta podría ser una oportunidad para vincular a los equipos de alto rendimiento nacionales con proyectos Big Data (Lee, 2014).

Energía. En lo que refiere a energía eléctrica, existe un relevamiento de la Compañía Administradora del Mercado Mayorista Eléctrico en tiempo real sobre consumo que podría servir para determinar patrones de comportamiento y predecir problemas en el sistema.

Entretenimiento. Redes sociales, juegos, aplicaciones móviles.

Nanotecnología. Dado el desarrollo del área en el país, se podría aplicar para la administración y el almacenamiento de datos, y brindar solución a las limitaciones del ancho de banda. Desde el Ministerio de Ciencia se están otorgando subsidios para proyectos de nanotecnología que incluyan nanomateriales, nanointermediarios y nanosensores.

Salud. El avance en la digitalización de historias clínicas e imágenes ofrece una oportunidad para lograr reducciones de costos, detección de irregularidades y eficiencia en los tratamientos médicos. Hay que tener en cuenta que el sector salud constituye un amplio espectro que involucra no solo a los prestadores médicos, sino también a la industria farmacéutica y de insumos de salud, sindicatos, obras sociales y otros actores.

SSI. Se desarrollan motores de búsqueda, herramientas para la búsqueda de información basada en contextos temáticos y se investiga en software de *data mining* genómico y farmacéutico.

Subsidios. Estrechamente vinculado al transporte, se cuenta con información de la tarjeta SUBE que, junto a otras bases de datos, permitiría una mejor asignación de los subsidios.

Transporte. A partir de herramientas georreferenciales se pueden realizar modelos de administración del tránsito, transporte público y planificación urbana. También se pueden realizar aplicaciones dinámicas continuas de información del transporte público tales como la detección de problemas, frecuencia y patrones de uso del servicio por parte de los pasajeros (tarjeta SUBE).

Es importante tener en cuenta que las aplicaciones de Big Data no se darán de forma automática. Su implementación no solo está sujeta a la disponibilidad de trabajadores calificados o la previa inversión en TIC, sino también a las barreras que puedan imponer el contexto institucional, político y cultural particular de los sectores, como así también los intereses económicos de los actores involucrados. Asimismo, un tema central en la aplicación de políticas exitosas será el tratamiento de la privacidad de los datos de las personas para que no haya un daño en su integridad y un descrédito de la herramienta que obstaculice un mejor aprovechamiento de sus beneficios. En este marco, la Argentina ha desarrollado capacidades técnicas y potencialidades que permitirían aprovechar este avance informacional-tecnológico en la contribución para generar un sendero de desarrollo económico y social virtuoso.

Apéndice. Técnicas y tecnologías vigentes para el tratamiento de Big Data

Actualmente, las técnicas y tecnologías para la captura, procesamiento, análisis y visualización de Big Data se clasifican en tres categorías de herramientas de procesamiento: por lotes, por flujos (*streaming*) y de análisis interactivo (Chen y Zhang, 2014). Las primeras se basan en la infraestructura de Apache Hadoop, tales como Mahout y Dryad. Para herramientas de procesamiento por flujos suelen utilizarse plataformas Storm o S4, mientras que para el análisis interactivo las plataformas suelen ser Dremel de Google y Apache Drill.

Técnicas

Las técnicas que se enuncian aquí son algunas de las usadas, por lo que esta lista no pretende ser exhaustiva; además, debe tenerse en cuenta que los desarrollos en esta materia están en permanente crecimiento.

A/B testing. Un grupo de control es comparado con una variedad de grupos de testeos para determinar cuál es el mejor procedimiento para alcanzar un objetivo particular. Big Data permite realizar varios testeos al asegurar grupos suficientemente grandes.

Association rule learning. Conjunto de técnicas para descubrir relaciones entre variables en grandes bases de datos.

Data mining. Técnicas para extraer patrones de grandes bases de datos al combinar métodos estadísticos y *machine learning*. Estas técnicas incluyen *association rule learning*, *cluster analysis*, clasificación y regresión.

Genetic algorithms. Técnica de optimización inspirada en procesos de evolución natural —supervivencia del más apto—. Usualmente son caracterizados como algoritmos evolutivos.

Machine learning. Vinculado al diseño y desarrollo de algoritmos que permiten a las computadoras desarrollar comportamientos basados en datos empíricos. *Natural language processing* es un ejemplo de *machine learning*.

Redes neuronales. Modelos computacionales inspirados en la estructura de redes neuronales biológicas —por ejemplo, células y conexiones dentro del cerebro— para buscar patrones en el *set* de datos y optimización.

Tecnologías

Cassandra. Administrador de base de datos de código abierto (*open source*) para el tratamiento de grandes cantidades de datos en un sistema distribuido. Es un proyecto Apache Software.

Nube de cómputo (o *cloud computing*). Paradigma computacional en el que los recursos de computación altamente escalables, a menudo configurados como un sistema distribuido, se ofrecen como un servicio a través de una red.

Extract, transform, and load. Herramientas de software utilizadas para extraer datos de fuentes externas, transformarlos para satisfacer las necesidades operativas y cargarlos en una base de datos.

Hadoop. Estructura de software de código abierto para el procesamiento de grandes bases de datos en sistemas distribuidos. Su desarrollo fue inspirado por Google's MapReduce y Google File System, desarrollado originalmente en Yahoo! y actualmente administrado como proyecto de Apache Software Foundation.

HBase. Base de datos abierta, distribuida y no relacionada, modelada en Google's Big Table. Originalmente desarrollada por Powerset, en la actualidad es administrada por Apache Software Foundation como parte de Hadoop.

MapReduce. Modelo de programación inspirado por Google para procesamiento de grandes bases de datos en sistemas distribuidos. Fue implementado también por Hadoop.

SQL (*structured query language* o lenguaje de consulta estructurado). Lenguaje de computación diseñado para administrar base de datos relacionales, permitiendo especificar distintos tipos de operaciones.

Visualización

Las herramientas de visualización permiten comunicar, entender y mejorar el análisis de los resultados obtenidos a partir de Big Data, en forma eficiente.

Tag Cloud. O nube de texto ponderada, donde las palabras utilizadas más frecuentemente aparecen en forma más grande que las menos frecuentes.

Clustergram. Técnica de visualización utilizada para mostrar cómo miembros de una base de datos se asignan a un *cluster* a medida que el número de *clusters* se incrementa.

History flow. Técnica de visualización que grafica la evolución de un documento a medida que este es editado por múltiples autores contribuyentes.

Spatial information flow. Muestra cómo fluyen los datos en forma espacial, es decir, de una región, país, ciudad o localidad a otra.

BIBLIOGRAFÍA

- ÁMBITO FINANCIERO (2014), "GLASER: 'Necesitamos una internet menos dependiente de EE.UU.'", versión *on line*, Buenos Aires, 4 de abril. Disponible en <<http://www.ambito.com/diario/noticia.asp?id=735458>>.
- ARMAH, NII AYI (2013), "Big Data Analysis: The Next Frontier", *Bank of Canada Review*, pp. 32-39. Disponible en <<http://www.bankofcanada.ca/wp-content/uploads/2013/08/boc-review-summer13-armah.pdf>>.
- BARLETTA, FLORENCIA, MARIANO PEREIRA, VERÓNICA ROBERT Y GABRIEL YOGUEL (2013), "Argentina: dinámica reciente del sector de software y servicios informáticos", *Revista CEPAL*, N° 110, agosto. Disponible en <<http://repositorio.cepal.org/handle/11362/11618>>.
- BARRANTES, ROXANA, VALERIA JORDÁN Y FERNANDO ROJAS (2013), "La evolución del paradigma digital en América Latina", en Jordán, Valeria, Hernán Galperín y Wilson Peres (coords.), *Banda ancha en América Latina: más allá de la conectividad*, Santiago, CEPAL/CE. Disponible en <http://www.cepal.org/es/publicaciones/35399-banda-ancha-america-latina-mas-alla-la-conectividad>
- BIEM *et al.* (2010), *Real-Time Traffic Information Management using Stream Computing*, Nueva York, IBM TJ Watson Research Center.
- BOLLIER, DAVID (2010), *The Promise and Peril of Big Data*, Washington, The Aspen Institute. Disponible en <http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf>.
- BOYD, DANAH Y KATE CRAWFORD (2012), "Critical questions for Big Data", *Information, Communication & Society*, vol. 15, N° 5, pp. 662-679. Disponible en <<http://dx.doi.org/10.1080/1369118X.2012.678878>>.
- BRYNJOLFSSON, ERIK, LORIN M. HITT Y HEEKYUNG H. KIM (2011), "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?", SSRN. Disponible en <http://www.a51.nl/storage/pdf/SSRN_id1819486.pdf>.

- CAMPANARIO, SEBASTIÁN (2013), "Big Data: ¿una oportunidad o un festival de humo de colores?", *La Nación*, 13 de octubre. Disponible en <<http://www.lanacion.com.ar/1628427-big-data-una-oportunidad-o-un-festival-de-humo-de-colores>>.
- CARVALHO *et al.* (2014), *A Smart Platform for Precision Agriculture*, Cisco White Paper.
- CESSI (2014), "Reporte semestral sobre el Sector de Software y Servicios Informáticos de la República Argentina", 1^{er} Semestre 2013, Observatorio Permanente de la Industria del Software y Servicios Informáticos de la República, Cámara de Empresas de Software y Servicios Informáticos. Disponible en <<http://www.cessi.org.ar/opssi>>.
- CHEN, C. L. PHILIP Y CHUN-YANG ZHANG (2014), "Data-Intensive Applications, Challenges, Techniques and Technologies: A Survey on Big Data", *Information Sciences*, N° 275, Macau (China), Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, pp. 314-347. Disponible en <<http://www.cs.unibo.it/~montesi/CBD/Articoli/SurvyBigData2.pdf>>.
- CHOI, DAE WON (2013), "Banda ancha y política industrial: la experiencia coreana", en Jordán, Valeria, Hernán Galperín y Wilson Peres (coords.), *Banda ancha en América Latina: más allá de la conectividad*, Santiago, CEPAL/CE. Disponible en <<http://www.cepal.org/es/publicaciones/35399-banda-ancha-en-america-latina-mas-alla-de-la-conectividad>>.
- CHOU, JUI-SHENG Y ABDI SURYADINATA TELAGA (2014), "Real-time detection of anomalous power consumption", *Renewable and Sustainable Energy Reviews*, vol. 33, mayo, pp. 400-411.
- COMISIÓN EUROPEA (2013), "Big Data. Analytics & Decision Making", Business Innovation Observatory, *case study* 8.
- CUMBLEY, RICHARD Y PETER CHURCH (2013), "Is 'Big Data' creepy?", *Computer Law & Security Review*, vol. 29, N° 5, Londres, pp. 601-609.
- DUMBILL, EDD (2012), "What is big data?", O'Reilly. Disponible en <<http://strata.oreilly.com/2012/01/what-is-big-data.html>>.
- EINAV, LIRAN Y JONATHAN LEVIN (2013), "The Data Revolution and Economic Analysis", NBER Innovation Policy and the Economy Conference, abril.
- FISHER, ANNE (2013), "Big Data could generate millions of new jobs", *Fortune*, mayo. Disponible en <<http://fortune.com/2013/05/21/big-data-could-generate-millions-of-new-jobs/>>.
- FORTE, MIGUEL ÁNGEL (2015), *Modernidad: tiempo, forma y sentido*, Buenos Aires, Eudeba.

- , SERGIO PIGNUOLI OCAMPO, SANTIAGO CALISE, MATÍAS PALACIOS Y MATÍAS ZITELLO (2012), "Las TIC como problemática de la teoría sociológica. Una aproximación conceptual a la comunicación digitalizada desde la teoría general de sistemas sociales autorreferenciales y autopoieticos", *Entramados y Perspectivas. Revista de la Carrera de Sociología*, vol. 2, N° 2, pp. 205-226. Disponible en <<http://publicaciones.sociales.uba.ar/index.php/entramadosyperspectivas/article/download/137/121>>.
- GARTNER (2011), "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". Disponible en <<http://www.gartner.com/newsroom/id/1731916>>.
- GORDON, ROBERT J. (2014), "The Demise of US Economic Growth: Restatement, Rebuttal, and Reflections", *Working Paper N° 19895*, National Bureau of Economic Research. Disponible en <<http://www.nber.org/papers/w19895.pdf>>.
- HARDING, MATTHEW Y MICHAEL LOVENHEIM (2013), "The Effect of Prices on Nutrition: Comparing the Impact of Product- and Nutrient-Specific Taxes", diciembre. Disponible: <http://gastronomiaycia.republica.com/wp-content/uploads/2014/10/estudio_consumo_bebidas_estados_unidos.pdf>.
- HARFORD, TIM (2014), "Big data: are we making a big mistake?", *FT Magazine*, 28 de marzo. Disponible en <<http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2zdXz8WBN>>.
- HILBERT, MARTIN (2013a), "Big Data for Development: From Information to Knowledge Societies", Washington, United Nations ECLAC. Disponible en <<http://ssrn.com/abstract=2205145>>.
- (2013b), "CEPAL Charlas Sobre Sistemas Complejos Sociales (CCSSCS)", video. Disponible en <<http://www.martinhilbert.net/CCSSCS.html>>.
- KELLY, JEFF (2014), "Big Data Vendor Revenue and Market Forecast", Wikibon. Disponible en <http://wikibon.org/wiki/v/Big_Data_Vendor_Revenue_and_Market_Forecast_2012-2017>.
- KRUGMAN, PAUL (2014), "Networks and Economic History", *The New York Times*, 13 de noviembre. Disponible en <<http://nyti.ms/112wYnT>>.
- LAZER, DAVID, RYAN KENNEDY, GARY KING Y ALESSANDRO VESPIGNANI (2014), "The Parable of Google Flu: Traps in Big Data Analysis", *Science Magazine*, vol. 343, N° 6176, 14 de marzo, pp. 1203-1205. Disponible en <<http://gking.harvard.edu/files/gking/files/0314policyforumff.pdf>>.

- LEE, DAVE (2014), "Big Data, el as bajo la manga de los equipos de fútbol del futuro", *La Nación*, 21 de abril. Disponible en <<http://www.lanacion.com.ar/1682716-big-data-el-as-bajo-la-manga-de-los-equipos-de-futbol-del-futuro>>.
- LI, HONGFEI, DHAIVAT PARIKH, QING HE, BUYUE QIAN, ZHIGUO LI, DONGPING FANG Y ARUN HAMPAPUR (2014), "Improving rail network velocity: A machine learning approach to predictive maintenance", *Transportation Research Part C: Emerging Technologies*, vol. 45, agosto, pp. 17-26.
- MANOVICH, LEV (2012), "Trending: The Promises and the Challenges of Big Social Data", en Gold, Matthew (ed.), *Debates in the Digital Humanities*, Minneapolis, The University of Minnesota Press, pp. 460-476.
- MCKINSEY (2011), "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute. Disponible en <http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation>.
- (2013), "The big-data revolution in US health care: Accelerating value and innovation", Center for US Health System Reform Business Technology Office. Disponible en <http://www.mckinsey.com/insights/health_systems_and_services/the_big-data_revolution_in_us_health_care>.
- OECD (2013), "Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by 'Big Data'", *OECD Digital Economy Papers*, N° 222, junio, OECD Publishing. Disponible en <<http://dx.doi.org/10.1787/5k47zw3fcp43-en>>.
- PERES, WILSON Y MARTN HILBERT (2010), "Information Societies in Latin America and the Caribbean Development of Technologies and Technologies for Development", ECLAC. Disponible en <http://www.cepal.org/publicaciones/xml/3/43803/Libro_Cepal_98.pdf>.
- PIKETTY, THOMAS Y EMMANUEL SAEZ (2003), "Income Inequality in the United States, 1913-1998", *The Quarterly Journal of Economics*, vol. 118, N° 1, pp. 1004-1042.
- RUBENS, PAUL (2014), "Un aliado tecnológico para evitar que la comida termine en la basura", *La Nación*, 19 de marzo. Disponible en <<http://www.lanacion.com.ar/1673266-un-aliado-tecnologico-para-evitar-que-la-comida-termine-en-la-basura>>.
- SCIDEV.NET (2014), "Big obstacles ahead for big data for development". Disponible en <<http://www.scidev.net/global/data/feature/obstacles-big-data-development.html>>.

SDSS-III (2008), "Massive Spectroscopic Surveys of the Distant Universe, the Milky Way Galaxy, and Extra-Solar Planetary Systems". Disponible en <<http://www.sdss3.org/collaboration/description.pdf>>.

TEKINER Y KEANE (2013), *Big Data Framework*, School of Computer Science, The University of Manchester, Manchester, UK.

THE NEW YORK TIMES (2013), "Is Big Data an Economic Big Dud?". Disponible en <<http://www.nytimes.com/2013/08/18/sunday-review/is-big-data-an-economic-big-dud.html?pagewanted=all&r=0>>.

UN Global Pulse (2012), "White paper: Big Data por Development: Opportunities & Challenges". Documento creado a partir de conversaciones mantenidas entre miembros de universidades de Estados Unidos (Massachusetts Institute of Technology, University of California at Berkeley, Stanford, entre otras) y empresas tecnológicas líderes (Yahoo!, Google, Microsoft, IBM, HP). Disponible en <<http://www.unglobalpulse.org/projects/BigDataforDevelopment>>.

UNITED NATIONS UNIVERSITY (2012), *Rainfall Variability, Food Security and Human Mobility: an approach for generating empirical evidence*, Bonn, United Nations University Institute for Environment and Human Security (UNU-EHS). Disponible en <<http://www.ehs.unu.edu/file/get/9921.pdf>>.

WIRED (2008), "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete". Disponible en <http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory>.

——— (2010), "Sergey Brin's Search for a Parkinson's Cure". Disponible en <http://www.wired.com/2010/06/ff_sergeys_search/all/>.

DT 3

FACUNDO MALVICINO es licenciado en Economía de la Universidad de Buenos Aires, magíster en Ciencia Política y Sociología por la Facultad Latinoamericana de Ciencias Sociales y docente en la Universidad Argentina de la Empresa y la Universidad Nacional de General Sarmiento (UNCS). Es investigador en el programa “Big Data para el desarrollo” del CIECTI.

GABRIEL YOGUEL es docente-investigador titular de la UNCS y coordinador del Área de Economía del Conocimiento de esa universidad. Forma parte del CIECTI y posee una muy importante trayectoria en investigación en temáticas de economía de la innovación y sistemas complejos. En ese campo es autor de numerosas publicaciones en revistas nacionales e internacionales y de varios libros.

