

# Grandes datos, grandes desafíos para las ciencias sociales

MARÍA ELENA MENESES ROCHA\*

*Resumen: El presente artículo analiza el debate actual sobre el denominado Big Data desde las ciencias sociales, enfatizando algunas discusiones de orden epistemológico sobre su eventual utilidad para el entendimiento de lo social. Se analiza la retórica alrededor de los datos como fuente de valor, así como algunas propuestas metodológicas para abordar problemáticas sociales mediante el análisis de dichos constructos sociales. Se sostiene que este novedoso giro cuantitativo exige un tratamiento ético por parte del científico social, una nueva práctica científica, pero sobre todo una mirada objetiva que permita conocer su utilidad para la comprensión de los fenómenos sociales.*

*Abstract: This paper analyzes the current debate on Big Data in the Social Sciences, emphasizing certain epistemological discussions about its possible utility for the understanding of the social. The rhetoric about data as a source of value is analyzed as well as certain methodological proposals to explore social problems through the analysis of these social constructs. It is argued that this novel quantitative twist requires ethical treatment on the part of social scientists, a new scientific practice, but above all an objective view that will reveal its usefulness in understanding social phenomena.*

*Palabras clave:* Big Data, epistemología, metodología, ciencias sociales, redes sociales.

*Keywords:* Big Data, epistemology, methodology, Social Sciences, social networks.

Un artículo publicado en la revista *Science* en 2011 revelaba que en 2020 habrá 50 veces más información que la generada en aquel año, de la cual 90% se encontrará desestructurada (Hilbert y López, 2011). Los incuestionables avances de la computación, la generalización de la información en redes y la penetración de dispositivos tecnológicos en la vida cotidiana dan como resultado ensamblajes de datos que son el insumo a través del cual se movilizan la economía y la sociabilidad.

\* Doctora en Ciencias Políticas y Sociales por la Universidad Nacional Autónoma de México. Tecnológico de Monterrey, Campus Ciudad de México. Temas de especialización: medios de comunicación, Internet y cultura política digital. Calle del Puente 222, Ejidos de Huipulco, Delegación Tlalpan, 14380, Ciudad de México.

A los estudiosos de los fenómenos sociales, el hecho de que existan más de 3 000 millones de individuos conectados a Internet y más de 7 000 millones de suscripciones a telefonía celular en el mundo les conduce a indagar la utilidad de la avalancha de información que se genera a partir del paradigma informacional para la mejor comprensión de sus objetos de estudio. Este fenómeno comienza a configurar nuevos campos disciplinarios, como la ciencia de datos, las ciencias sociales computacionales e incluso otras iniciativas disciplinarias, como las humanidades digitales.<sup>1</sup> Se podría señalar que la tecnología es parte fundamental de este conjunto de propuestas que tienen como eje a las ciencias computacionales cuyo saber es condición indispensable (Berry, 2012).

Big Data, cuya traducción al idioma español podría ser grandes cantidades de datos o macrodatos, es parte de la agenda de gobiernos y de empresas, y es menester de las ciencias sociales entender qué fenómenos sociales pueden ser comprendidos a través de éstos, qué dilemas éticos traen consigo y cómo su análisis puede ayudar a entender, proyectar y resolver problemáticas sociales.<sup>2</sup> Más allá de los discursos entusiastas que aluden a una “revolución de los datos” —usualmente provenientes de gobiernos y empresas colectoras de éstos—, resulta una tarea irrenunciable para los científicos sociales comprender de forma realista y crítica la transformación de datos en conocimiento útil para la sociedad. Esta tarea implica no solamente el diálogo con las ciencias computacionales, sino adquirir nuevos conocimientos, habilidades y lenguajes.

El lenguaje de la programación modifica las estrategias de investigación y las formas de trabajo, lo cual no es tarea fácil para los campos disciplinarios que se configuran desde un corpus teórico y métodos específicos (Benoit y Cukier, 2015). Se podría decir que a partir de los grandes datos presenciamos un momento en el cual se abren posibilidades para la imaginación sociológica, como la entendió Charles Wright Mills (2003), ya que amplía las posibilidades de observación, experimentación, y abre la puerta a una nueva practicidad. El propósito de este artículo es analizar el debate sobre los macrodatos desde las ciencias sociales, poniendo un

<sup>1</sup> De acuerdo con datos de la Unión Internacional de Telecomunicaciones, UIT (2015).

<sup>2</sup> Hasta el año 2016 la Real Academia de la Lengua Española no reconoce el concepto “Big Data”. Sobre los extranjerismos, señala que su tratamiento se debe hacer, en la medida de lo posible, apegándose a los rasgos gráficos y morfológicos del español. La literatura en español traduce el vocablo como macrodatos o grandes volúmenes de éstos.

particular énfasis en aquellos que se generan en las redes sociales digitales, para comprender las principales discusiones epistemológicas.

Las fuentes de datos tienen una relación directa con la digitalización. Los celulares, las imágenes satelitales y las redes sociales digitales son las tres principales fuentes, seguidas de aquellos procedentes de sensores y de las transacciones de los sectores público y privado, respectivamente (Ballivian, 2016).

Para dimensionar el volumen, variedad y velocidad de datos que se recogen en las redes sociales digitales, basta analizar las siguientes impresionantes cifras: en 2016 Facebook tenía 1 090 millones de usuarios activos diariamente, de los cuales 84% estaban fuera de Estados Unidos (Socialbakers, 2016). YouTube contaba con 1 000 millones de usuarios, registraba 4 billones de visitas diarias, en las cuales los usuarios subían 300 horas de video por minuto (DMR, 2016). Twitter, a su vez, tenía 310 millones de usuarios activos al mes, que en conjunto en ese año enviaron 500 millones de tuits diariamente (Smith, 2016).

Detrás de la sociabilidad que promueven en su retórica mercadológica, las redes sociales digitales constituyen un complejo entramado tecnológico, económico y político que les permite transformar la narrativa personal, los gustos y los estados de ánimo en datos, los cuales son su motor económico y base de su modelo de negocio. A pesar de la brecha digital que deja fuera a 51.8% de los habitantes del mundo, la vida personal y laboral, así como la vida política, transcurre en buena medida en las redes sociales digitales.

#### ENTRE LA RETÓRICA Y LA UTILIDAD PARA LA COMPRENSIÓN DE LO SOCIAL

Intentar comprender la ontología de los grandes datos no es objeto del presente artículo; en cambio, se pretende ofrecer algunas claves de su eventual utilidad para el entendimiento de lo social. Big Data es un término que comienza a generalizarse a partir del año 2008 y se encuentra en plena construcción conceptual para las ciencias sociales. Francis Diebold (2012) atribuye al científico de Silicon Graphics John Mashey haberlo usado por vez primera en 1999 para referirse al análisis de grandes bases de datos (Kitchin, 2014: 66). Años más tarde, diarios financieros como *The Economist*, *The Financial Times* y revistas como *Science* lo retomarían, contribuyendo con ello a su generalización. Se podría decir que aún ca-

rece de claridad conceptual, su definición varía dependiendo del campo que lo define, y hasta ahora predomina un enfoque económico y técnico.

Para comenzar a construir el concepto es necesario recurrir a los orígenes del vocablo *dato*, del latín *dare*, que significa dar. Lo que se entiende como dato, comúnmente asociado con información y conocimiento, ha cambiado a lo largo del tiempo. Daniel Rosenberg (2013: 27) ubica la aparición del término *data* en inglés en el siglo XVII, al que le procede un periodo de latencia para concluir que es hasta el siglo XX cuando alcanza un pico estadístico. Una búsqueda similar en español podría arrojar información distinta y probablemente su generalización puede enmarcarse en otra temporalidad. Si se indaga el volumen de libros con el vocablo Big Data en inglés con la aplicación Ngram Viewer de Google, su pico estadístico se localiza en el año 2000, lo que contribuye a darnos una idea de que su generalización es relativamente reciente.

A los datos se les adjudica en la actualidad ser fuente de valor económico y político, más si se trata de grandes volúmenes. A su retórica han contribuido empresas, organismos mundiales y gobiernos del mundo. Desde el movimiento de datos abiertos, al cual se han adherido paulatinamente los países democráticos, se considera a los datos como un bien público que favorece la transparencia y mejora la toma de decisiones políticas.<sup>3</sup> Para la Organización para la Cooperación el Desarrollo Económico (OCDE) (2013), al ser insumo de la economía digital, el tratamiento y análisis de los datos permite el desarrollo de nuevas industrias y productos, por lo que constituyen un nuevo modelo de toma de decisiones.

De acuerdo con un entusiasta estudio de la consultora McKinsey de 2013 (Manyika *et al.*, 2013), los datos abiertos ayudarían a crear valor por tres trillones de dólares al año en la economía global; crearán nuevas empresas y servicios, y “empoderarán” a los ciudadanos. Ese mismo año Neelie Kroes (2013), la vicepresidenta de la Comisión para la Agenda Digital Europea, señalaba que 1.7 millones de billones de bytes de datos

<sup>3</sup> La intención de las iniciativas de datos abiertos promovidas por organismos mundiales, Estados Unidos y la Unión Europea desde la década de los años noventa, es que los datos obtenidos por el sector público y organismos mundiales estén en la red a la disposición de cualquiera. Datos provenientes de censos, estadísticas nacionales, regionales y mundiales deben estar al alcance de los ciudadanos con el objeto de promover la transparencia y la rendición de cuentas. Un análisis crítico de este movimiento obliga a pensar en la asequibilidad de los recursos para acceder a ellos y traducirlos en conocimiento. La presidencia de Donald Trump en Estados Unidos, hasta ahora nación líder del movimiento de datos y gobierno abierto, obliga a tener reservas sobre su futuro.

se generaban por minuto, lo que equivale a más de seis megabytes por día por cada persona en el planeta. Si el conocimiento es el motor de la economía, decía la funcionaria europea, los datos son el combustible. Un ejemplo de este valor es el uso de grandes datos en la mercadotecnia para comprender el comportamiento del consumidor, hipersegmentar los mercados y dirigir campañas publicitarias personalizadas, lo cual convierte a los grandes datos en insumo imprescindible del sistema económico.

Los macrodatos han sido vinculados también al desarrollo. Un claro ejemplo son los proyectos que impulsa la Organización de Naciones Unidas bajo el nombre Global Pulse, los cuales demuestran cómo el análisis de grandes volúmenes de datos puede ayudar a frenar una epidemia como el ébola al mapear el recorrido de los casos, mediante teléfonos celulares, alertando a los servicios de salud de los diversos continentes. Gracias a la geolocalización de datos también se puede focalizar la ayuda en casos de desastres o amenazas, entre otras causas humanitarias.<sup>4</sup>

En tiempos recientes, algunas investigaciones globales se vinculan necesariamente a las ciencias sociales, como el estimado de flujos migratorios mediante las búsquedas en Google que ayuda a complementar las estadísticas nacionales y a trazar políticas públicas (UN Global Pulse, 2014). Los tuits también han permitido a la ONU, en colaboración con la universidad de Leiden, conocer las diferencias de acuerdo con el género en discusiones sobre el desarrollo global en temas relacionados con violencia, nutrición, cambio climático y la utilización de los recursos naturales (UN Global Pulse, 2016). Sin ser los únicos, sino apenas un par de ejemplos entre muchos otros, ayudan a tener una idea de lo que el análisis con Big Data puede aportar al conocimiento de fenómenos sociales complejos para mejorar la toma de decisiones.

En tanto que la práctica científica ha sido ya modificada con el análisis de cuantiosos datos en diversos campos disciplinarios, como la física y la biología, en las ciencias sociales comienzan a adoptarse (Heiberger y Riebling, 2016).

Poco a poco las huellas humanas embebidas en satélites *scanners* y redes sociales digitales comienzan a ser cuantificadas y analizadas por científicos sociales.

<sup>4</sup> Los proyectos de ciencias de datos de la ONU pueden consultarse en <http://www.unglobalpulse.org/>.

La energía social en Twitter, por ejemplo, se ha analizado con el objeto de comprender fenómenos sociopolíticos, ya que a través del análisis de millones de tuits se devela con contundencia empírica cómo se organiza la protesta contemporánea (Agarwal *et al.*, 2014; Bennett *et al.*, 2014).

Gracias al análisis de grandes datos se cuenta con evidencias sobre los efectos de la propaganda *online* en campañas políticas (Bond *et al.*, 2012). Otros fenómenos, como el contagio emocional, indagado a través de los datos provenientes de Facebook, resultan necesarios y útiles para comprender la política en la actualidad, que en buena medida ocurre en las redes sociales digitales (Fowler y Christakis, 2008; Kramer *et al.*, 2014). Prácticas digitales emergentes evidenciadas luego del análisis con grandes datos que inciden en la vida democrática como los *bots* (Bessi y Ferrara, 2016) y las noticias falsas (Allcott y Gentzkow, 2017) invitan a elaborar nuevas preguntas e hipótesis desde la sociología política.

En estudios económicos es posible realizar proyecciones bursátiles y detectar tasas de inflación en tiempo real rastreando los precios de millones de productos en el mundo, como sucede en el proyecto The Billion Prices Project del Tecnológico de Massachusetts (Cavallo y Rigobon, 2016). Estudios como el de Toole *et al.* (2015), que a través de los registros de llamadas por celular observó la movilidad y cambios de hábitos de los usuarios, para posteriormente identificar los efectos de despidos laborales en dos países europeos, son muestra de la utilidad de los macrodatos para comprender fenómenos como el desempleo.

A su vez, la gestión de las grandes ciudades en la actualidad difícilmente se lleva a cabo sin el análisis de las huellas que dejan sus habitantes en sus trayectos y consumo cultural (Townsend, 2013).

La práctica política tampoco está ajena a los datos masivos. En su campaña para la reelección presidencial de 2012, Barack Obama utilizó el análisis de Big Data para personalizar su campaña ante más de 69 millones de estadounidenses reunidos en bases de datos (Issenberg, 2012). En 2016, de acuerdo con el diario *The Washington Post*, Hillary Clinton usó la analítica de datos mediante un algoritmo para guiar su estrategia de campaña (Wagner, 2016).

El periodismo está inmerso en el inmenso caudal de datos y da lugar a un nuevo giro cuantitativo en su práctica que se ha dado en llamar periodismo de datos. Los *Papeles de Panamá*, que develaron una compleja red de evasión de impuestos a escala global, son un ejemplo de cómo el poder informático ayudó a descomprimir y analizar 11.5 millones de archivos con más de un millón de imágenes, más de dos millones

de archivos en PDF, tres millones de bases de datos y cinco millones de correos electrónicos en la más voluminosa filtración de la historia del periodismo (Meneses, 2016). En tanto, para el derecho, los datos personales embebidos en dispositivos tecnológicos y redes son objeto de protección mediante regulaciones *ex profeso* al relacionarse con el derecho fundamental a la privacidad.

### ¿QUÉ TAN *BIG*?

Hay diferentes tipos de datos y metadatos en bruto (datos de los datos), pues no son iguales los datos que se obtienen de la investigación genómica a los que se coleccionan de sensores para medir el consumo de energía o de una red social digital como Twitter. Éstos preceden a la información y el conocimiento que requiere métodos y técnicas de procesamiento especializado de acuerdo con sus especificidades (Kitchin, 2014; Weinberger, 2012).

Describir y analizar los tipos de datos y los desafíos que representan es necesario, aunque no es el propósito de este artículo. Basta decir que no todos los datos son iguales, ni en su infraestructura, ni en los tipos de ensamblaje, ni en su valor. Tampoco los retos y dilemas éticos para el investigador son los mismos. Para definir Big Data frecuentemente se alude a ciertos atributos técnicos, como el volumen, que se refiere a la cantidad de datos disponibles que por su tamaño requieren sistemas computacionales sofisticados para capturarlos y por su difícil almacenaje resulta imposible guardarlos en una computadora común.

A mediados del año 2012, en una encuesta de la empresa IBM, se preguntó a expertos el volumen de datos requerido para ser considerado Big Data, a lo que respondieron que para ser considerada como tal la base de datos tiene que ser superior a 1 114 terabytes (Gandomi y Haider, 2015). Para darse una idea, 16 millones de fotografías de Facebook pueden ser almacenadas en un terabyte; sin embargo, estas apreciaciones varían. La literatura disponible de carácter técnico coincide en señalar, además del volumen, otros dos atributos: variedad y velocidad. Un ensamblaje de datos cumple con el criterio de variedad si los metadatos son suficientemente diversos para permitir el análisis en distintos niveles.

Si tomamos una base de datos de Twitter, en un tuit veremos que además del texto de 140 caracteres hay ligas a una serie de recursos culturales diversos, ya sean videos, audios o fotografías, los cuales la mayoría de las

veces se encuentran desestructurados temporal y espacialmente. Podemos observar todas sus convenciones comunicativas como los tuits, retuits, menciones y hashtags cuantificados. Podemos saber cuántos seguidores tiene un usuario y a cuántos sigue, cuántos tuits ha escrito y desde cuándo es usuario. Un cruce de variables de tuits puede arrojar correlaciones interesantes sobre el fenómeno que se quiera explicar.

Un tercer atributo técnico es la velocidad, rasgo que se aprecia en el continuo e imparable flujo de la generación de los datos. Basta pensar en el número de individuos que pasan los controles de seguridad en los aeropuertos del mundo, los registros de las cámaras de vigilancia de las grandes ciudades, las transacciones bancarias y el impresionante volumen de datos que genera un usuario de un teléfono celular con acceso a Internet, cuyo registro obra en poder de las empresas de telecomunicaciones, las cuales en algunos países son obligadas a conservarlos por periodos de uno a dos años por razones de seguridad y combate al crimen.<sup>5</sup>

De esta forma, las tres V —volumen, variedad y velocidad— son los atributos técnicos que pueden ayudar a definir Big Data y sobre los cuales hay cierto consenso en la literatura técnica y mercadológica que existe al respecto. Sin embargo, para las ciencias sociales estos tres atributos, que pueden observarse en la tabla 1, son insuficientes, ya que solamente aluden al tamaño descomunal de datos y algunas especificidades relacionadas, cuando existen otras propiedades necesarias para entenderlo, como la veracidad, objetividad, representatividad y los dilemas éticos asociados.

Big Data no es la única forma de ensamblaje de datos, sólo es una más; por ello, como proponen los investigadores Kitchin y McArdle (2016), es imprescindible diferenciarla de otros conglomerados de datos que no son considerados Big Data, a los que la literatura especializada denomina Small Data, cuya captura ocurre de manera controlada y responden a un diseño estadístico y conceptual *ex profeso*.

<sup>5</sup> La Ley Federal de Telecomunicaciones y Radiodifusión de México obliga a las empresas de telecomunicaciones y de contenidos audiovisuales a retener los datos de los usuarios por dos años.



TABLA 1  
LOS TRES ATRIBUTOS TÉCNICOS DE BIG DATA

<i>Datos</i>	<i>Small Data</i> (pequeños volúmenes)	<i>Big Data</i> (grandes volúmenes)
Volumen	Limitado	Casi ilimitado
Velocidad	Lenta	Rápido y continuo
Variedad	Limitado	Amplia

FUENTE: Elaboración propia con base en Kitchin y McArdle (2016).

Su volumen es limitado, se capturan en tiempos largos y tienen escasa variedad, usualmente limitada a sus reactivos. Un ejemplo de pequeños datos son las encuestas que han servido como insumo para las ciencias sociales desde hace décadas. En estos casos las tres variables —volumen, variedad y velocidad— no se cumplen a cabalidad de acuerdo con los atributos técnicos. En contraste, Big Data aparece ante el investigador desordenado y caótico, por lo cual éste tiene que limpiarlo y ordenarlo para que le permita observar variables con el fin de convertir esa estructura de datos en categorías de conocimiento a ser interpretadas.

## LOS DATOS COMO CONSTRUCCIÓN SOCIAL

Lo que podemos denominar “condición de Big Data”, dada por el tamaño y la velocidad, es una simple descripción de las bases de datos y conduce a ampliar la mirada hacia un debate epistemológico sobre las posibles consecuencias de una nueva forma de empirismo o, como algunos entusiastas sugieren, del fin de la teoría en favor del poder algorítmico (Anderson, 2008). De acuerdo con estas posturas, que han sido objeto de diversas críticas desde las ciencias sociales, el análisis con macrodatos mediante técnicas analíticas, como el aprendizaje automatizado (Machine Learning) o la minería de datos (Data Mining), permite encontrar patrones con los que se pueden elaborar modelos útiles para el pleno entendimiento de lo social.

Estas posturas son matizadas por los científicos sociales que consideran que los grandes ensamblajes representan un acercamiento más de las ciencias sociales a las naturales, como ha llegado a sugerir Bruno

Latour (2007, 2011); de esta forma, el nuevo giro cuantitativo constituye un nuevo intento en esa dirección. En este mismo sentido, Dominique Boullier (2016), investigador del Laboratorio de Ciencias Políticas de la Universidad de París, ha llegado a comparar la ola de Big Data con otros momentos de la sociología y su dependencia de las encuestas para explicar lo social. La ola de los datos masivos sería una más en la búsqueda de las ciencias sociales por contar con evidencias de la vida social.

Más que anunciar la muerte de la teoría, los ensamblajes de datos masivos permiten probar hipótesis diversas y llegar a niveles de focalización imposibles para un sondeo o encuesta. Un ejemplo de esta focalización quirúrgica la podemos observar en los estudios de audiencias. Netflix es un ejemplo de cómo los gustos de sus usuarios convertidos en datos se traducen en una rigurosa personalización que le permite a la empresa satisfacer preferencias personales, lo cual supera en mucho a las mediciones de ratings de la era análoga. Como sugieren Viktor Mayer-Schönberger y Kenneth Cukier (2014) en uno de los libros más citados sobre Big Data, el muestreo estadístico propio de tiempos de escasez de información estaría pasando a la historia cuando ésta es abundante, un supuesto desplazamiento que requiere realizar más estudios antes de darlo como un hecho.

Ni la muerte de la teoría, ni la supuesta comprensión plena y universal de la sociedad, como señalan los entusiastas, convencen a los científicos sociales, ya que medir patrones no significa poder explicarlos, para lo cual la teoría es indispensable.

Más allá de las tres V, se puede sostener que los datos no existen fuera de las ideas, de los instrumentos, las prácticas y el contexto que enmarca su creación e interpretación (Gitelman, 2013). Argumentos como éstos son los que predominan en las posturas críticas desde las ciencias sociales, para las cuales todo dato es una representación de un fenómeno y, por tanto, una construcción social en la que intervienen diversas mediaciones (Kitchin, 2014; Boullier, 2016; Shoemaker y Reese, 2014). Incluso hay quienes han considerado el concepto un oxímoron, por su ambigüedad y contradicción (Gitelman y Jackson, 2013; Bowker, 2014).

Otras aristas asociadas a los datos preocupan a algunos científicos sociales, quienes consideran que existe la posibilidad de que sean utilizados como una forma de control social, un sistema de vigilancia construido a partir de nuestros hábitos (Poster, 1996), en tanto que otros los estudian como práctica política porque reflejan los valores que construyen los sujetos políticos sobre lo público (Johnson, 2014).

Para las ciencias sociales los datos no son neutrales, sino una construcción en la que ocurren diversas mediaciones hasta llegar al laboratorio, las cuales todo investigador, independientemente del campo disciplinario, debe tomar en cuenta.

### ¿OBJETIVIDAD Y NEUTRALIDAD?

La discusión en las ciencias sociales que se puede consultar en revistas especializadas de reciente creación como *Big Data and Society* se ha centrado en el análisis de la supuesta objetividad de los datos para el análisis de lo social y en el valor que los grandes datos aportan a la investigación de fenómenos sociales. La relevancia de este tipo de ensamblaje para las ciencias sociales no estaría dada por atributos relacionados con el tamaño y especificidades técnicas, sino por la forma en que se generan, en las relaciones con otros datos, por las conexiones de un individuo, de un individuo con otros, de grupos sociales, o bien, de la propia información.

Para Boyd y Crawford (2012) los datos no son neutrales, sino que expresan las intenciones de quienes los diseñan. En uno de los artículos más citados en la literatura disponible proponen como provocación intelectual para las ciencias sociales indagar de manera realista y crítica las implicaciones de lo que puede ser considerado un nuevo sistema de conocimiento. Para las investigadoras, Big Data es un fenómeno cultural, tecnológico y mitológico; este último atributo dada el “aura de verdad” que acompaña a su retórica.

Los algoritmos con que se escarban las huellas más profundas de los datos no siempre son neutrales. En 2016, *ProPublica*, el medio de investigación periodística estadounidense, analizó los algoritmos utilizados por el sistema judicial para predecir los casos de reincidencia delictiva. Encontró que los algoritmos habían sido creados con sesgos racistas (Angwin *et al.*, 2016). Evidencias como ésta llevan a concluir que las nociones como la objetividad y la representatividad de lo social, así como la brecha digital que se profundiza con la datocracia, son temas insoslayables para los científicos sociales, que no pueden aceptar sin cuestionamientos la retórica de que los grandes datos nos acercan a un universo probabilístico donde el comportamiento humano puede predecirse para construir una “mejor sociedad” (Crawford, Miltner y Gray, 2014).

Los datos, en su mayoría, pertenecen a los gobiernos o a unas cuantas empresas tecnológicas localizadas fundamentalmente en Estados Unidos.

En el caso de los gobiernos, el movimiento de Datos Abiertos comienza a permear en algunos países democráticos; sin embargo, y paradójicamente, quienes usan los datos para su beneficio económico son las empresas que los resguardan celosamente.

La recolección de datos y las múltiples transacciones que realizan empresas como Google, Facebook, Amazon, Apple, así como las compañías de telecomunicaciones —en algunos casos—, no se llevan a cabo mediante contratos transparentes sobre su utilización y protección. Esta situación es observable en las redes sociales digitales, cuyos usuarios aceptan condiciones de uso de sus datos sin conocer a cabalidad lo que implican. Entre desconocimiento, tecnicismos y prácticas opacas se aceptan condiciones sin conocer, en la mayoría de los casos, las consecuencias.

El diseño tecnológico aparece como una forma de opacidad de las redes sociales digitales que, bajo el halo de la sociabilidad, son en realidad empresas colectoras de datos. Facebook, por ejemplo, reduce las expresiones a sus siete iconos; Twitter, por su lado, permite frases de 140 caracteres como máximo; se trata entonces de decisiones preestablecidas para cuantificar la sociabilidad (Van Dijck, 2013). Estas empresas, que constituyen una de las más importantes infraestructuras de datos del mundo, dejan poco espacio para que los usuarios tengan pleno control sobre sus datos personales y no comparten con los gobiernos toda la información que guardan por su supuesto compromiso con la libertad de expresión, y mucho menos con la comunidad científica.

Los dilemas éticos para el investigador son un tema que no puede soslayarse. Capturar una base de datos de una red social digital, además de que debe atenderse a la cantidad de datos que las empresas permiten, conlleva tener en ese ensamblaje datos de millones de personas. A veces se trata de datos personales delicados que deben ser manejados con riguroso cuidado para no afectar la privacidad de las personas, por lo cual resulta urgente debatir y definir parámetros éticos para la investigación con datos masivos.

La concentración del conocimiento para acceder a los datos, procesarlos y analizarlos es también obstáculo y desafío (Gurstein, 2011). Éste no es un argumento menor, ya que la falta de datos y el conocimiento especializado se encuentran focalizados en ciertas universidades estado-

unidenses y europeas, en contraste con una muy discreta producción de regiones como América Latina.

Paradójicamente, América Latina, en el caso de las redes sociales digitales, es la región que más las utiliza en el mundo (Cepal, 2015: 63). La Cepal señala que es en esta región donde se encuentra el uso más intenso de las redes sociales, con una participación de 78.4% de los usuarios de Internet en 2013. El organismo atribuye este fenómeno a los bajos costos de hardware y a la globalización de patrones de consumo, así como a la débil oferta latinoamericana de redes sociales digitales y buscadores.<sup>6</sup> Esta participación en la generación de datos, paradójicamente, no se traduce en más investigación regional. El algoritmo de Google Académico encuentra 75 900 textos sobre Big Data del año 2000 al año 2016, en contraste con 3 510 en español en el mismo lapso.

La región latinoamericana es una gran fuente de datos para las empresas de la economía digital, que los blinda en sus cajas negras para reformular sus negocios. Aunque el algoritmo de Google no necesariamente ofrece toda la investigación que se realiza en español, es un indicador útil para explicar esta brecha del conocimiento. Un claro ejemplo de esta inequidad y falta de parámetros éticos lo ofrece un reciente caso protagonizado por Facebook y la Universidad de Cornell, quienes manipularon el algoritmo de la red social para observar las reacciones psicológicas de los usuarios ante informaciones negativas. Los resultados son relevantes y útiles, pues muestran que las expresiones de otros influyen en nuestras emociones, con lo cual se fortalecen las hipótesis relativas a que el contagio emocional no es sólo un fenómeno que se desprende de la interacción física, sino también de la virtual. Sin embargo, para obtener esta evidencia Facebook introdujo en el perfil de 700 000 personas determinados contenidos, lo que causó una gran controversia en el mundo académico (Kramer *et al.*, 2014). El caso puso en la mesa de discusión el tema de la ética en la investigación apoyada por las empresas tecnológicas, lo cual vuelve necesaria la elaboración de guías de buenas prácticas para los científicos sociales que trabajan con Big Data, como las que se discuten en las Naciones Unidas y en la National Science Foundation de Estados Unidos, y las cuales pueden ser emuladas en otros contextos (Global Pulse, 2017; Fiske y Hauser, 2014).

<sup>6</sup> En 2014 México contaba con sólo 21 millones de visitantes al mes a videos en línea, los cuales ocupaban en promedio 6.1 minutos por video, lo cual rebasa el promedio mundial de 4.8 minutos.

Otras miradas analíticas se detienen en la invasión a la privacidad como una consecuencia que debe ser indagada por las ciencias sociales (Zuboff, 2015; Crawford, Miltner y Gray, 2014; Poster, 1996). Otros enfoques orientan su mirada a cuestionar si Big Data y sus técnicas analíticas pueden arrojar hallazgos más fiables que los obtenidos a través de herramientas de corte cualitativo. La investigadora Rebecca Lemov (2014) recuerda que desde hace un siglo la sociología y la antropología capturan bases de datos sobre la experiencia humana en ensamblajes diversos. La historiadora sugiere que la pregunta a responder por las humanidades y las ciencias sociales es si analizar con Big Data es más relevante para el conocimiento que hacerlo con aquellos datos recogidos mediante técnicas cualitativas. Una provocación que habrá de ser validada mediante la indagación desde nuestras disciplinas.

En cuanto a la forma en que los datos son recogidos en las redes sociales digitales, Ellison, Heino y Gibbs (2006) sostienen que deben ser observados por el investigador como imaginaciones, opiniones e ideas que presentan sólo ciertos aspectos que son seleccionados para la construcción de la personalidad en línea. Si bien hay datos valiosos para comprender la textura social, también hay situaciones que se escapan a las técnicas de minería de datos o al aprendizaje automatizado y a la interpretación que hacen los investigadores que trabajan con estos voluminosos ensamblajes.

Los macrodatos obligan a un análisis político y ético, ya que por las condiciones en que los datos fueron generados, almacenados, capturados, minados y analizados, no permiten que toda la experiencia de la vida cotidiana pueda ser captada y analizada a partir de ellos. ¿En dónde quedan los pequeños actores?, se preguntan con razón los investigadores Nick Couldry y Allison Powell (2014) en un intento por reivindicar el valor de los acercamientos cualitativos. Otros investigadores, como Lev Manovich (2012), que ha realizado estudios en redes sociales visuales como Instagram, no se detienen en los alcances epistemológicos ni en la subjetividad y prefieren enfocarse en la novedad de las variables que se pueden obtener sin que ocurra un desplazamiento de las técnicas tradicionales de las ciencias sociales, como lo son las técnicas etnográficas.

De acuerdo con este enfoque más conciliador, los grandes datos tienen la ventaja de que las ciencias computacionales pueden capturar huellas, movimientos, opiniones y prácticas culturales de millones de personas, lo cual es imposible para el etnógrafo tradicional y para los estudios demoscópicos. Se trata, como ha sugerido Bruno Latour (2007), de una acumulación gigantesca de huellas e imaginaciones embebidas en el

hardware y software que pueden develar hechos y situaciones únicas, así como favorecer nuevas preguntas y correlaciones, lo cual constituye una gran oportunidad para la reinención de las ciencias sociales.

En la discusión son abordadas otras problemáticas más allá de la neutralidad, como los obstáculos y dificultades para realizar indagaciones socioculturales con grandes ensamblajes de datos. Por ejemplo, la falta de entrenamiento de los científicos sociales puede traducirse en una claudicación ante las ciencias computacionales en favor de un empirismo que desplace las categorías de pensamiento de las ciencias sociales.

En mayo de 2016 la editorial Sage encuestó a 9412 científicos sociales de todo el mundo para conocer si realizaban investigación con Big Data. La encuesta reveló que 33% sí había estado involucrado en investigaciones con datos masivos, y que 49% planeaba hacerlo. De los que ya realizan indagaciones con éstos, 60% dijo haberlo hecho en los últimos 12 meses. Sobre las dificultades, 42% mencionaron los fondos y 32% el acceso a los datos.<sup>7</sup> Del total de los que respondieron, 3 302 fueron académicos estadounidenses, cuatro veces más que los que respondieron en Reino Unido y 46 más que los que respondieron en México (Metzler *et al.*, 2016). Más allá de su fiabilidad, este estudio arroja cierta evidencia con respecto a la centralidad de Estados Unidos en la generación de conocimiento sobre el tema.

En un escenario optimista, los límites disciplinarios para abordar problemas complejos se podrían traducir en una colaboración multidisciplinaria. Estas nuevas formas de generación de conocimiento requieren una distribución geográfica del trabajo, basada en plataformas de colaboración que requieren de la comunicad científica nuevas habilidades para la atracción de fondos concurrentes y que son rasgo de la práctica científica en la actualidad, lo que Ankeny y Leonelli (2016) denominan los nuevos repertorios del cambio científico: una especie de ensamblaje de conceptos, materialidades, estrategias, procedimientos y políticas institucionales que son imprescindibles para la ciencia.

Se puede sostener que se trata de una discusión abierta en la ciencia, toda vez que las mediaciones que ocurren en todas las etapas de la cadena de valor de los grandes datos, desde la generación hasta el análisis, no son una preocupación exclusiva de los científicos sociales. Sabina Leonelli, la investigadora de ciencias de la vida (London School of Economics

<sup>7</sup> La encuesta se puede consultar en <<https://us.sagepub.com/sites/default/files/compsocsci.pdf>>.

and Political Science, 2015), considera que indagar las condiciones de recolección de las primeras evidencias antes de integrarlas en las bases de datos es de suma relevancia para la biología experimental.

Es así que de la rigurosidad en cada etapa de la cadena de valor de los datos, que va desde la generación hasta el análisis, depende la fiabilidad de los estudios, sostiene la investigadora (figura 1). La recolección, almacenaje, dispersión, visualización, así como los factores tecnopolíticos, pueden incidir en la investigación científica, la cual debe realizar un análisis situado de los datos. A decir de la investigadora, analizar las formas de recolección, curación y almacenaje, como parte de un contexto tecnopolítico, aparece como un imperativo de este nuevo reto del quehacer científico.

FIGURA 1  
CADENA DE VALOR DE LOS DATOS



FUENTE: OCDE (2013). *Supporting Investment in Knowledge Capital, Growth and Innovation*.

Si bien el debate no ha concluido, hay consenso en que el análisis de grandes volúmenes de datos puede tener un valor para el conocimiento de lo social, y para lo cual es imprescindible que los científicos sociales realicen sus propios experimentos, no sin antes reconocer las condiciones sociotécnicas en que los datos fueron generados. Esta advertencia resulta útil para aquellos investigadores que analizan datos provenientes de redes sociales digitales.

Para analizar Facebook, Instagram, YouTube o Twitter se debe partir del reconocimiento no sólo de las especificidades sociotécnicas de dichas redes, sino también de la falta de representatividad, de la subjetividad de los usuarios al postear y compartir una información, de los atributos de tipo geopolítico, así como de su innegable poder económico y centralidad en la cultura de nuestro tiempo.

Los datos, como se ha sostenido, son construcciones sociales en las cuales interviene una multiplicidad de mediaciones en cada una de las fases de la cadena de valor. A éstas se añade la subjetividad del usuario de redes y aplicaciones, el cual, si bien está fuertemente condicionado, encuentra espacios para la agencia individual y colectiva con implica-



ciones difíciles de prever, como ha sucedido en Egipto, Túnez, Madrid, Nueva York y en otros contextos donde la indignación colectiva se ha articulado a través de la comunicación en las redes.

## EL DISCURSO DEL MÉTODO DIGITAL

Ante la innegable *datificación* de la sociedad y la búsqueda de métodos y técnicas para la comprensión de fenómenos socioculturales, se han realizado propuestas basadas en técnicas computarizadas que pueden capturar y analizar datos. Sin que sean las únicas propuestas, es pertinente mencionar las iniciativas metodológicas de Bruno Latour para las ciencias sociales y la de Franco Moretti para las humanidades.

Latour, en su Laboratorio de Datos de la Facultad de Ciencias Políticas de la Universidad de París, puso en marcha su metodología de las controversias: un conjunto de técnicas para observar, analizar y mapear fenómenos sociales con la ayuda de programas informáticos. Se trata de una versión didáctica, como él mismo lo asume, de su Teoría del Actor-Red (TAR). En las humanidades, la pulsión por incorporar el saber computacional a los estudios literarios fue materializada desde 2011 por el italiano Franco Moretti en el Laboratorio de Literatura de la Universidad de Stanford. El investigador impulsa desde hace más de una década una corriente de estudios literarios cuantitativos mediante técnicas computarizadas, así como el análisis de redes sociales (ARS) con software de visualización.

Su propuesta metodológica, a la que denomina lectura distante (Moretti, 2013), contraria a la lectura cercana —la del análisis literario convencional cuya unidad de análisis es un texto— es tomada como pionera para quienes hoy se desenvuelven dentro del movimiento de las humanidades digitales. En sus estudios incorpora miles de novelas en bases de datos y ha causado hasta la fecha intensos debates sobre la cuantificación de los estudios literarios que se proponen fundamentalmente identificar patrones y articular modelos. Moretti le llama Literatura-mundo, en alusión a la propuesta teórica de Sistema-mundo desarrollada por el sociólogo Immanuel Wallerstein.

En los últimos años han proliferado en el mundo laboratorios ligados a las ciencias sociales y a las humanidades en los que se desarrollan experimentos con datos masivos como los antes expuestos, pero también con datos pequeños. Con datos masivos se pretende encontrar patrones

y, en su caso, elaborar modelos y proyectar fenómenos sociales mediante técnicas analíticas específicas, entre las que destacan, como he señalado, el aprendizaje automatizado y la minería de datos, de las cuales se obtienen correlaciones e incluso proyecciones.

La proyección es un tema controversial en la discusión sobre el valor de los grandes datos para el conocimiento de lo social, ya que los fenómenos sociales son susceptibles a la contingencia y a una infinidad de variables que pueden salirse de control, razón por la cual sus posibilidades de predicción son puestas en duda por diversos investigadores. El investigador de la Universidad de California Matthew Hilbert (2015) señala que todo dato pertenece al pasado, por lo cual sugiere que el científico social debe reconocer que todo pronóstico social es falible. Sin embargo, en lo que parece una discusión inacabada, los análisis en tiempo real y proyectivos, respectivamente, han demostrado ser relevantes para la economía al permitir la medición de conductas y reacciones al instante, así como la simulación de escenarios que resulta útil para la toma de decisiones.

Estas dudas que aún se ciernen sobre la utilidad de los estudios con macrodatos pueden aminorarse si en vez de descartar éstos *a priori* se realizan estudios utilizándolos para después compararlos con los que son producto derivado de otros ensamblajes que permiten palpar sus aportaciones y limitaciones de forma objetiva. Comparar y explicar se antoja necesario para conocer los beneficios de cada constructo, así como para considerar un eventual uso complementario.

Los estudios de ciencias genómicas y las proyecciones epidemiológicas y de morbilidad son ejemplos de cómo la analítica de grandes datos en algunos casos puede proyectar situaciones.

## PROPUESTAS ANALÍTICAS CON SMALL DATA

Además de los análisis con macrodatos existen otras propuestas analíticas más cercanas al enfoque cualitativo a las que se han acercado sociólogos, comunicólogos y antropólogos en tiempos recientes, para las cuales el volumen de datos no es condición, a diferencia de las técnicas computarizadas y el procesamiento de datos, considerando que el objeto de estudio de estos investigadores son los objetos digitales y el proceso de producción cibercultural.

Los denominados métodos digitales son una apuesta para investigar a la sociedad y sus prácticas culturales con un grupo de técnicas “nativas

de la Red” (Rogers, 2013). La propuesta desarrollada por Richard Rogers, el investigador de la universidad de Ámsterdam, consiste en tomar a la WWW como laboratorio, para lo cual se desarrollan programas computacionales específicos. Los más comunes son las arañas de la web (*web crawlers*) y las aplicaciones para realizar curaduría de datos provenientes de diversas plataformas digitales de forma rápida.

Los métodos digitales son de gran ayuda para el investigador y en ocasiones resultan útiles para seleccionar el corpus de un estudio en cualquier plataforma digital con criterios menos subjetivos. Éstos funcionan como prototipos, están en constante proceso de mejora y son de código abierto. Tienen como limitante estar supeditados a las posibilidades de las plataformas. Si se quiere indagar en YouTube, el investigador se tiene que atener a las métricas que brinda esta empresa de Google; si lo que se desea es crear una herramienta de *scraping* de Facebook, estará circunscrita a lo que esta red social posibilita.

Sin embargo, los métodos digitales escarban en minutos millones de datos, los cuales a un científico social le tomaría meses obtener. Por ejemplo, para indagar algún fenómeno cultural en YouTube existe una herramienta que permite extraer las URL de cientos de miles de videos rápidamente. Ese corpus requerirá que el investigador realice la limpieza correspondiente conforme a las preguntas que quiera responder, que tenga capacidad de procesamiento y de análisis, y finalmente que utilice las aplicaciones de visualización para poder observar las correlaciones provenientes del caudal de datos.

Dentro de este marco de propuestas para las cuales el volumen de datos no es condición única, se encuentra la analítica cultural (Manovich, 2012), la etnografía digital (Pink, *et al.*, 2016) o la antropología digital (Horst y Miller, 2012), las cuales trabajan con volúmenes pequeños de datos.

Para estas propuestas la Red, como el lugar donde se gesta la cultura en forma de bytes y píxeles, es objeto y método, y ambos coinciden en la necesidad de profundizar, adaptar y mezclar las técnicas etnográficas tradicionales para observar ciertos fenómenos culturales con la ayuda del saber computacional.

El grupo de iniciativas que trabajan con Small Data estarían más ligadas a las denominadas humanidades digitales, en tanto que las investigaciones con Big Data lo estarían al campo emergente de las ciencias sociales computacionales, que bien pueden ser consideradas un nuevo giro positivista (Kitchin, 2014).

Este caudal de propuestas experimentales con datos masivos y pequeños volúmenes de datos influye cada vez más en las facultades de ciencias sociales y humanidades del mundo, aunque su producción aún es modesta como para poder evaluar su utilidad para el conocimiento de los fenómenos sociales y culturales.

Por lo pronto, en un escenario de recursos limitados para la investigación, todas estas propuestas e iniciativas surgen como oportunidad de captación de fondos, variable que no puede dejar de analizarse ante la intención de las universidades por subirse a la ola de los métodos digitales y Big Data.

Se trata de una convergencia de campos y metodologías diferentes a los de la etapa analógica, que privilegia la preminencia del saber computacional como vehículo para la comprensión de fenómenos sociales y que exige una mirada multidisciplinaria (Berry, 2012). Aunque este artículo no pretende profundizar en los cambios que el saber computacional ha traído a la práctica científica, podría considerarse que favorece nuevas formas de construcción del conocimiento con los obstáculos que acarrea el trabajo entre disciplinas.

Además del saber computacional, otras disciplinas cobran relevancia en este escenario científico. Todo dato debe ser representado para ser comprendido, así que, paralelamente a los análisis de Big Data y con métodos digitales, destaca una, cada vez más vasta, producción de programas para la visualización de datos, que pueden ser definidos y comprendidos como representaciones de los datos, y cuyo diseño obliga a la colaboración multidisciplinaria.<sup>8</sup> Se trata de innovaciones técnicas e incluso artísticas para visualizar datos y extraer correlaciones para el entendimiento de fenómenos sociales.

Hay quienes consideran a la analítica visual como un área de investigación que cobra cada vez mayor relevancia y que tiene por objetivo hacer visible la complejidad del descomunal volumen de información (Keim *et al.*, 2008). Mapas interactivos, como los georreferenciados, los de sentimientos o de intenciones de voto presentes en las campañas políticas de la actualidad, son ejemplo de esta novedosa forma

<sup>8</sup> Algunos, como Gephi y CartoDB, se encuentran entre los más comunes, aunque no son los únicos. Gephi es un programa desarrollado en 2009 por estudiantes de la Universidad de Campiègne, Francia. La herramienta es usada por académicos y medios de comunicación para visualizar la estructura de las redes sociales. CartoDB es un programa desarrollado en 2007 por los españoles Javier de la Torre y Sergio Álvarez para visualizar datos en mapas.

de socialización del conocimiento, producto de la investigación con datos masivos y con datos no masivos. Son comunes las visualizaciones de redes provenientes de investigaciones de la denominada ciencia de redes, a partir de la cual una buena cantidad de investigadores analizan y contribuyen a entender fenómenos sociopolíticos y culturales como las movilizaciones sociales de la actualidad, las noticias falsas o el discurso de odio en la Red.<sup>9</sup>

## REFLEXIONES FINALES

Las iniciativas analizadas en este artículo pueden considerarse una dirección científica novedosa (Conte *et al.*, 2012) que produce una abundante cantidad de artículos científicos, lo cual se observa en las revistas especializadas y congresos dedicados a exponer avances y discutir sobre este giro científico, que está permeando cada vez más las ciencias sociales. A pesar de esto aún no cuenta con el consenso de la comunidad de científicos sociales para ser considerado un nuevo paradigma, de acuerdo con el pensamiento de Thomas Kuhn sobre las revoluciones científicas.

Que cambie el método y que, una vez explorado el corpus, las preguntas y las hipótesis se enuncien mediante algoritmos no es suficiente. Es necesario que este grupo de técnicas novedosas, mediadas por el saber computacional, den como resultado nuevos paradigmas teóricos. No se observa, pues, hasta ahora, un desplazamiento, sino una convivencia de formas de abordaje empírico de los fenómenos sociales, con las cuales los científicos sociales habrán de tomar distancia con el discurso técnico y mercadológico, teniendo en cuenta que los mecanismos causales de toda investigación podrían ser fortalecidos con nuevos datos producto de los procesos tecnológicos y económicos analizados en el presente artículo.

Probablemente no es cuestión de inventar métodos, como señala Boullier (2016), sino de crear nuevos procesos de reflexividad para las

<sup>9</sup> El estudio de redes complejas puede inscribirse en las denominadas Ciencias de Redes, a las que el Consejo Nacional de Investigación de Estados Unidos define como las que estudian las representaciones de redes biológicas o sociales que dan lugar a modelos de predicción. El análisis de redes sociales (ARS) basado en este cuerpo de conocimiento multidisciplinario es utilizado frecuentemente por estudiosos de las redes sociales digitales.

sociedades, para lo cual son necesarios diversos tipos de análisis, incluidos los que se realizan con pequeños o grandes datos.

Frente a la avalancha de estas formas de representación de lo social en bases de datos, y ante las tentaciones de dar por muerta la teoría y privilegiar el método, es menester de las ciencias sociales no perder de vista lo esencial, y recurrir a la advertencia que hacen Craig Calhoun y Michel Wieviorka (2013) en su *Manifiesto por las Ciencias Sociales* cuando señalan que quienes fetichizan las técnicas “corren el riesgo de pasar de largo lo esencial: el contenido intelectual de su aportación...”.

El reto es que entender los nuevos métodos y técnicas no equivalga a dejar de comprender los fenómenos sociales que el avance tecnológico trae consigo. Hacer nuestras preguntas, bajo nuestras categorías, así como encontrar otras nuevas, parece un camino necesario para no quedar avasallados por el saber computacional. Se trata de poner a la tecnología al servicio del conocimiento de lo social. Mientras tanto, no podemos asumir *a priori* que los grandes datos—objeto de análisis del presente artículo— constituyen una simple retórica sin antes indagar en sus posibilidades y limitaciones para el conocimiento de la sociedad.

## BIBLIOGRAFÍA

- AGARWAL, Sheetal, Lance Bennett, Courtney Johnson y Shawn Walker (2014). “A model of crowd enabled organization: Theory and methods for understanding the role of Twitter in the occupy protests”. *International Journal of Communication* 8: 646–672.
- ALLCOTT, Hunt, y Matthew Gentzkow (2017). “Social media and fake news in the 2016 election”. *Journal of Economic Perspectives* 31 (2): 211–239.
- ANDERSON, Chris (2008). “The end of theory: The data deluge makes the scientific method obsolete” [en línea]. *Wired*. Disponible en <<http://www.wired.com/2008/06/pb-theory/>> [última consulta: 2 de junio de 2016].
- ANGWIN, Julia, Jeff Larson, Surya Mattu y Lauren Kirchner (2016). “Machine Bias” [en línea]. *ProPublica*. Disponible en <<https://googl/8MAfhK>> [última consulta: 1 de marzo de 2016].

- ANKENY, Rachel, y Sabina Leonelli (2016). "Repertoires: A post-Kuhnian perspective on scientific change and collaborative research". *Studies in History and Philosophy of Science Part A* 60: 18-28.
- BALLIVIAN, Amparo (2016). "El uso de Big Data para estadísticas oficiales y los Objetivos de Desarrollo Sostenible". Ponencia presentada en el Seminario La Revolución del Big Data en Estudios Sociales, El Colegio de México, México, 29 de abril.
- BENNETT, Lance, Alexandra Segerberg y Shawn Walker (2014). "Organization in the crowd: peer production in large-scale networked protests". *Information, Communication and Society* 17 (2): 232-260.
- BENOIT, Kenneth, y Kenneth Cukier (2015). "The challenges of Big Data to Social Sciences" [en línea]. Disponible en <<http://www.lse.ac.uk/website-archive/newsAndMedia/videoAndAudio/channels/publicLecturesAndEvents/player.aspx?id=2888>> [última consulta: 1 de marzo de 2015].
- BERRY, David (editor) (2012). *Understanding Digital Humanities*. Londres: Palgrave MacMillan.
- BESSI, Amy, y Emilio Ferrara (2016). "Social bots distort the 2016 US presidential election online discussion" [en línea]. *First Monday*. Disponible en <<http://firstmonday.org/ojs/index.php/fm/article/view/7090/5653a>> [última consulta: 29 de diciembre de 2016].
- BOND, Robert, Christopher Fariss, Jason Jones, Adam Kramer, Cameron Marlow, Jaime Settle y James Fowler (2012). "A 61-million-person experiment in social influence and political mobilization". *Nature* 489 (7415): 295-298.
- BOULLIER, Dominique (2016). "Big Data challenges for the social sciences: from society and opinion to replications" [en línea]. Disponible en <<https://arxiv.org/abs/1607.05034>> [última consulta: 5 de marzo de 2015].
- BOWKER, Geoffrey (2014). "The theory/data thing". *International Journal of Communication* 8 (5): 1795-1799.
- BOYD, Danah, y Kate Crawford (2012). "Critical questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon". *Information, Communication and Society* 15 (5): 662-679.

- BRUNS, Axel, y Jean Burgess (2012). "Researching news discussion on Twitter: New methodologies". *Journalism Studies* 13 (5-6): 801-814.
- CALHOUN, Craig, y Michel Wieviorka (2013). "Manifiesto por las ciencias sociales". *Revista Mexicana de Ciencias Políticas y Sociales* 58 (217): 29-60.
- CAVALLO, Alberto y Roberto Rigobon (2016). "The Billion Prices Project: Using online prices for measurement and research". *The Journal of Economic Perspectives* 30 (2): 151-178.
- COMISIÓN ECONÓMICA PARA AMÉRICA LATINA Y EL CARIBE (CEPAL) (2015). "La nueva revolución digital: de la Internet del consumo a la Internet de la producción" [en línea]. Disponible en <[http://repositorio.cepal.org/bitstream/handle/11362/38604/S1600780\\_es.pdf?sequence=4&isAllowed=y](http://repositorio.cepal.org/bitstream/handle/11362/38604/S1600780_es.pdf?sequence=4&isAllowed=y)> [última consulta: 17 de septiembre de 2016].
- CONTE, Rosaria, Nigel Gilbert, Giulia Bonelli, Claudio Cioffi-Revilla, Guillaume Deffuant, Janos Kertesz, Vittorio Loreto, Suzy Moat, J-P Nadal, Anxo Sanchez, Andezej Nowak, Andreas Flache, Maxi San Miguel y Dirk Helbing (2012). "Manifesto of computational social science". *The European Physical Journal Special Topics* 214 (1): 325-346.
- CORASANITI, Nick, y Azam Ahmed (2016). "Donald Trump to visit Mexico after more than a year of mocking it" [en línea]. *The New York Times*. Disponible en <[http://www.nytimes.com/2016/08/31/us/politics/donald-trump-mexico-enrique-pena-nieto.html?ref=nyt-esy\\_r=0](http://www.nytimes.com/2016/08/31/us/politics/donald-trump-mexico-enrique-pena-nieto.html?ref=nyt-esy_r=0)> [última consulta: 1 de septiembre de 2016].
- COULDRY, Nick, y Alison Powell (2014). "Big Data from the bottom up" [en línea]. *Big Data & Society* 1 (2). Disponible en <<https://doi.org/10.1177/2053951714539277>>
- CRAWFORD, Kate, Kate Miltner y Mary Gray (2014). "Critiquing Big Data: Politics, ethics, epistemology. Special section introduction". *International Journal of Communication* 8 (10): 1663-1672.
- DIEBOLD, Francis (2012). "On the origins and development of Big Data: the phenomenon, the term, and the discipline" [en línea]. Disponible en <<https://economics.sas.upenn.edu/sites/economics.sas.upenn.edu/files/12-037.pdf>> [última consulta: 16 de marzo de 2016].



- DMR (2016). “145 Amazing YouTube Statistics” [en línea]. DMR. Disponible en <<http://expandedramblings.com/index.php/youtube-statistics/>> [última consulta: 16 de marzo de 2016].
- ELLISON, Nicole, Rebecca Heino y Jennifer Gibbs (2006). “Managing impressions online: Self-presentation processes in the online dating environment”. *Journal of Computer-Mediated Communication* 11 (2): 415-441.
- FISKE, Susan, y Robert Hauser (2014). “Protecting human research participants in the age of Big Data”. *Proc Natl Acad Sci* 111 (38): 13675–13676.
- FOWLER, James y Nicholas Christakis (2008). “Estimating peer effects on health in social networks: a response to Cohen-Cole and Fletcher; Trogdon, Nonnemaker, Pais”. *Journal of Health Economics* 27 (5): 1400-1405.
- GANDOMI, Amir, y Murtaza Haider (2015). “Beyond the hype: Big Data concepts, methods, and analytics”. *International Journal of Information Management* 35 (2): 137-144.
- GITELMAN, Lisa (editor) (2013). “*Raw Data*” is an Oxymoron. Cambridge: MIT Press.
- GITELMAN, Lisa, y Virginia Jackson (2013). “Introduction”. En *Raw Data is an Oxymoron*, 1-14. Cambridge: MIT Press.
- GLOBAL PULSE (2017). “Big Data for development and humanitarian action: Towards responsible governance” [en línea]. Disponible en <[http://unglobalpulse.org/sites/default/files/Big\\_Data\\_for\\_Development\\_and\\_Humanitarian\\_Action\\_Report\\_Final\\_0.pdf](http://unglobalpulse.org/sites/default/files/Big_Data_for_Development_and_Humanitarian_Action_Report_Final_0.pdf)> [última consulta: 27 de junio de 2017].
- GURSTEIN, Michael (2011). “Open data: Empowering the empowered or effective data use for everyone?” [en línea]. *First Monday* 16 (2). Disponible en <<http://firstmonday.org/article/view/3316/2764>> [última consulta: 5 de noviembre de 2016].

- HEIBERGER, Raphael, y Jan Riebling (2016). "Installing computational social science: Facing the challenges of new information and communication technologies in social science" [en línea]. *Methodological Innovations* 9: 1-11. Disponible en <<https://doi.org/10.1177/2059799115622763>>
- HILBERT, Martin (2015). "The theory, practice and limits of Big Data for the social sciences" [en línea]. *Opening Speech at UC Davis Institute for Social Sciences*. Disponible en <<http://www.martinhilbert.net/the-theory-practice-and-limits-of-big-data-for-the-social-sciences>> [última consulta: 1 de marzo de 2016].
- HILBERT, Martin, y Priscila López (2011). "The world's technological capacity to store, communicate, and compute information" [en línea]. *Science*. Disponible en <<http://science.sciencemag.org/content/332/6025/60>> [última consulta: 1 de marzo de 2016].
- HORST, Heather, y Daniel Miller (2012). "The digital and the human: A prospectus for digital anthropology". En *Digital Anthropology*, editado por Heather Horst y Daniel Miller, 3-35. Londres: Berg Publication.
- INTERNET WORLD STATISTICS (2016). *Internet Usage Statistics* [en línea]. Disponible en <<http://www.internetworldstats.com/stats.htm>> [última consulta: 10 de marzo de 2016].
- ISSENBERG, Sasha (2012). "How Obama's team used Big Data to rally voters" [en línea]. *MIT Technology Review*. Disponible en <<https://www.technologyreview.com/s/509026/how-obamas-team-used-big-data-to-rally-voters>> [última consulta: 10 de marzo de 2016].
- JOHNSON, Jeffrey Alan (2014). "From open data to information justice". *Ethics and Information Technology* 16 (4): 263-274.
- KEIM, Daniel, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer y Guy Melançon (2008). "Visual analytics: Definition, process, and challenges" [en línea]. *Information Visualization*. Disponible en <<https://hal-lirmm.ccsd.cnrs.fr/lirmm-00272779/document>> [última consulta: 10 de marzo de 2016].
- KITCHIN, Rob (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures and their Consequences*. Los Ángeles: Sage.

- KITCHIN, Rob, y Gavin McArdle (2016). "What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets" [en línea]. *Big Data and Society* 3 (1). Disponible en <<https://doi.org/10.1177/2053951716631130>> [última consulta: 12 de marzo de 2016].
- KRAMER, Adam, Jamie Guillory y Jeffrey Hancock (2014). "Experimental evidence of massive-scale emotional contagion through social networks". *Proceedings of the National Academy of Sciences* 111 (24): 8788-8790.
- KROES, Neelie (2013). "The economic and social benefits of Big Data" [en línea]. *European Commission*. Disponible en <[http://europa.eu/rapid/press-release\\_SPEECH-13-450\\_en.htm](http://europa.eu/rapid/press-release_SPEECH-13-450_en.htm)> [última consulta: 15 de junio de 2016].
- LATOUR, Bruno (2007). "Beware, your imagination leaves digital traces". *Times Higher Literary Supplement* 6 (4): 129-131.
- LATOUR, Bruno (2011). *Mapping Controversies* [en línea]. Disponible en <<http://www.bruno-latour.fr/node/362>> [última consulta: 1 de marzo de 2016].
- LEMOV, Rebecca (2014). "Collecting, organizing, trading Big Data" [en línea]. *Swiss STS meeting*. Disponible en <<https://wp.unil.ch/stsbigdata/rebecca-lemov-2/>> [última consulta: 21 de abril de 2016].
- LEONELLI, Sabina, y Mark Carrigan (2015) "Sabina Leonelli: What constitutes trustworthy data changes across time and space" [en línea]. *lse: Impact of Social Sciences Blog*. Disponible en <<http://blogs.lse.ac.uk/impactofsocialsciences/2015/01/19/philosophy-of-data-science-series-sabina-leonelli/>> [última consulta: 11 de abril de 2016].
- MANOVICH, Lev (2012). "Trending. The promises and the challenges of Big Data science" [en línea]. *Debates on Digital Humanities*. Disponible en <<https://files.digitalmethods.net/readings/dmisummer11reader.pdf>> [última consulta: 3 de abril de 2016].
- MANYIKA, James, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh y Angela Hung (2013). "Big Data: The next frontier for innovation, competition, and productivity" [en línea]. *McKinsey Global Institute*. Disponible en <<http://www.mckinsey.com/business-functions/>>

- digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation> [última consulta: 11 de abril de 2016].
- MAYER-SCHÖNBERGER, Viktor, y Kenneth Cukier (2014). *Big Data: A Revolution that will Transform how we Live, Work, and Think*. Nueva York: Houghton Mifflin Harcourt.
- MENESES, María Elena (2016). “#Panamapapers. El resurgimiento del periodismo de investigación”. *Foreign Affairs Latinoamérica* 16 (3): 104-110.
- METZLER, Katie, David Kim, Nick Allum y Angella Denman (2016). “Who is doing computational social science? Trends in Big Data research” (White paper) [en línea]. Londres: Sage. Disponible en <<https://us.sagepub.com/sites/default/files/CompSocSci.pdf>> [última consulta: 11 de septiembre de 2016].
- MORETTI, Franco (2013). *Distant Reading*. Londres: Verso.
- ORGANIZACIÓN PARA LA COOPERACIÓN Y EL DESARROLLO ECONÓMICOS (OCDE) (2013). *Supporting Investment in Knowledge Capital, Growth and Innovation*. París: OCDE Publishing.
- PINK, Sarah; Heather Horst, John Postill, Larissa Hjorth, Tania Lewis y Jo Tacchi (2016). *Digital Ethnography. Principles and Practices*. Londres: Sage.
- POSTER, Mark (1996). “Databases as Discourse; or, Electronic Interpellations Mark Poster. Computers”. En *Computers, Surveillance and Privacy*, editado por David Lyon y Elia Zureik, 175-192. Minneapolis: University of Minnesota Press.
- RAINIE, Lee, y Barry Wellman (2012). *Networked: The New Social Operating System*. Cambridge: MIT Press.
- REAL ACADEMIA ESPAÑOLA (RAE). “Tratamiento de los extranjerismos” [en línea]. Disponible en <<https://goo.gl/wvkGHC>> [última consulta: 1 de febrero de 2016].
- ROGERS, Richard (2013). *Digital Methods*. Cambridge: The MIT Press.
- ROSENBERG, Daniel (2013). “Data before the fact”. En *Raw Data is an Oxymoron*, editado por Lisa Gitelman, 15-41. Cambridge: MIT Press.

- SHOEMAKER, Pamela, y Stephen Reese (2014). *Mediating the Message in the 21st Century*. Nueva York: Routledge.
- SMITH, Kit (2016). “44 Twitter Statistics for 2016” [en línea]. *Brandwatch*. Disponible en <<https://www.brandwatch.com/2016/05/44-twitter-stats-2016/>> [última consulta: 1 de febrero de 2016].
- SOCIALBAKERS (2016). “Facebook statistics directory” [en línea]. Disponible en <<https://www.socialbakers.com/statistics/facebook/>> [última consulta: 1 de diciembre de 2016].
- TOOLE, Jameson, Yu-Ru Lin, Erich Muehlegger, Daniel Shoag, Marta González y David Lazer (2015). “Tracking employment shocks using mobile phone data” [en línea]. *Journal of The Royal Society Interface* 12 (107). Disponible en <<http://rsif.royalsocietypublishing.org/content/royinterface/12/107/20150185.full.pdf>>
- TOWNSEND, Anthony (2013). *Smart Cities. Big Data, Civic Hackers and the Quest for a New Utopia*. Nueva York: Norton.
- UN GLOBAL PULSE (2014). “Estimating migration flows using online search data” [en línea]. *Global Pulse Project Series*. Disponible en <[http://www.unglobalpulse.org/sites/default/files/UNGP\\_ProjectSeries\\_Search\\_Migration\\_2014\\_0.pdf](http://www.unglobalpulse.org/sites/default/files/UNGP_ProjectSeries_Search_Migration_2014_0.pdf)> [última consulta: 12 de diciembre de 2016].
- UN GLOBAL PULSE (2016). “Sex disaggregation of social media posts” [en línea]. *Tool Series*. Disponible en: <[http://unglobalpulse.org/sites/default/files/2-pager%20Sex-Disaggregation%20-%20final\\_1.pdf](http://unglobalpulse.org/sites/default/files/2-pager%20Sex-Disaggregation%20-%20final_1.pdf)> [consulta: 29 de diciembre de 2016].
- VAN DIJCK, José (2013). *The Culture of Connectivity: A Critical History of Social Media*. Oxford: University Press.
- WAGNER, John (2016). “Clinton’s data-driven campaign relied heavily on an algorithm named Ada. What didn’t she see?” [en línea]. *The Washington Post*. Disponible en <<https://www.washingtonpost.com/news/post-politics/wp/2016/11/09/clintons-data-driven-campaign-relied-heavily-on-an-algorithm-named-ada-what-didnt-she-see/>> [última consulta: 12 de diciembre de 2016].

- WEINBERGER, David (2012). *Too Big to Know*. Nueva York: Basic Books.
- WRIGHT MILLS, Charles (2003). *La imaginación sociológica*. México: Fondo de Cultura Económica.
- ZUBOFF, Shoshana (2015). “Big other: Surveillance capitalism and the prospects of an information civilization”. *Journal of Information Technology* 30 (1): 75-89.

Recibido: 23 de enero de 2017

Aceptado: 9 de agosto de 2017