CIENCIAS SOCIALES Y *BIG DATA*

Representaciones políticas, disputas comunicacionales y política internacional

ESTEBAN ACTIS,
MARIANA BERDONDINI
Y SEBASTIÁN CASTRO ROJAS
[Compiladores]

CENTRO DE ESTUDIOS COMPARADOS
UNIVERSIDAD NACIONAL DE ROSARIO

UNR EDITORA

Ciencias Sociales y Big Data: . / Esteban Actis ... [et al.] ; compilado por Esteban Actis; Mariana Berdondini; Sebastián Ramiro Castro Rojas. - 1a ed. - Rosario: UNR Editora. Editorial de la Universidad Nacional de Rosario, 2021.
272 p.; 23 x 16 cm.

ISBN 978-987-702-432-6

1. Comunicación Política. I. Actis, Esteban, comp. II. Berdondini, Mariana, comp. III. Castro Rojas, Sebastián Ramiro, comp.

CDD 320.014

EQUIPO EDITOR

Directora UNR Editora
Nadia Amalevi

Coordinación
Nicolás Manzi

Diseño y maquetación
Joaquina Parma
Cristina Rosenberg

Corrección
Tomás Boasso
Ezequiel Hazan

©Esteban Actis
Universidad Nacional de Rosario, 2020.
Queda hecho el depósito que marca la Ley N° 11.723.
Ninguna parte de esta obra puede ser reproducida sin el permiso expreso del editor.
Impreso en Argentina.

UNR editora
Editorial de la Universidad Nacional de Rosario
Secretaría de Extensión Universitaria
Urquiza 2050 - S2000AOB / Rosario, República Argentina
www.unreditora.unr.edu.ar / editora@sede.unr.edu.ar

capítulo 3

El *Big Data* bajo la lupa: notas sobre el retrato de una época

Irene Lis Gindin
[Consejo Nacional de Investigaciones Científicas y Técnicas
Universidad Nacional de Rosario, Argentina]

Mariana Patricia Busso
[Consejo Nacional de Investigaciones Científicas y Técnicas
Universidad Nacional de Rosario, Argentina]

1. Introducción

Las utopías, deseos y realidades en torno al *Big Data* y su utilización en las ciencias sociales, vienen ocupando espacio en las reflexiones tanto académicas como periodísticas. El presupuesto según el cual los datos presentes en Internet, debido a su pretendido libre y fácil acceso, se encuentran disponibles para todo aquel que se interese en su compilación y abordaje debe, cuanto menos, ser puesto bajo la lupa. Así como la llegada de Internet trajo nuevas preocupaciones y modificó los modos de abordaje de nuestros objetos de estudio; así las ciencias sociales se encuentran aún en una etapa en donde la cautela sobre el uso de los macrodatos[1] debiera ser la regla, mas no la excepción. Si consideramos, a la par de Verón (2013), que la importancia de la WWW no se encuentra tanto en la última W (*Web*) sino en las dos primeras (*World Wide*), se comprende que "el emergente es, entonces, el alcance, el campo de aplicación, y no el concepto" (p. 278). Si hay, por tanto, una revolución en el acceso, convendría preguntarnos, ¿acceso a qué y de parte de quiénes?

Este capítulo se propone como objetivo dar cuenta de las principales reflexiones en torno a las problemáticas epistemológicas y metodológicas que surgen del abordaje de objetos de estudio que conllevan la utilización de

1 Usaremos el término "macrodato" como sinónimo de *Big Data*.

Big Data. Ese recorrido implicará poner en cuestión el presupuesto de que *necesariamente* toda investigación social actual debe, para ser considerada relevante, trabajar con grandes cantidades de datos, ya sea para describir, explicar o prever un determinado fenómeno. Por su parte, y para las investigaciones que efectivamente necesitan del *Big Data* para su realización, recuperaremos los desarrollos teóricos del área, con el fin de dar cuenta de ciertas problemáticas ligadas a esos trabajos. La veracidad, la objetividad, la neutralidad, la representatividad y los dilemas éticos (Meneses Rocha, 2018) se presentan como ejes a tener en cuenta.

Disyuntivas relacionadas a la cuestión de lo cualitativo y lo cuantitativo vuelven a aparecer a la hora de dar cuenta de estas problemáticas, en lo que pareciera ser no sólo una actualización de tal discusión, sino la manifestación de la fuerza misma de verdad adjudicada a la cantidad de datos. En efecto, el "sentido común científico", si se nos permite, pareciera asumir que estos datos –neutrales *per se*– son la garantía de una investigación exitosa, dejando de lado aspectos que consideramos insoslayables: la generación misma de los datos, las relaciones que éstos establecen entre sí, el contexto en el que son producidos, entre otras variables.

Por último, la guarda de esta enorme cantidad de material digital, en especial aquél que es extraído de los sitios de redes sociales en Internet, comienza a configurar una nueva forma de archivo que también merece ser estudiada. ¿Se trata de una mera agregación de información confusa o, por el contrario, de una colección de materiales para la que es necesario desplegar particulares estrategias y operaciones de accesibilidad y resguardo? ¿Estamos entonces en una época donde los metadatos se configuran como su nuevo archivo, y por ende, como legado a las generaciones futuras? A estas cuestiones nos dedicaremos a lo largo de este capítulo.

2. *Big Data*, sentidos y problemáticas asociadas

Empecemos por dar cuenta de a qué nos referimos cuando hablamos de *Big Data*. Una forma de ingreso a esta pregunta está dada por la propia definición del "dato". Puschmann y Burguess (2014a) realizan un interesante recorrido en torno a las acepciones de este término, con el fin de comprender la forma en la que emerge y se estabiliza una determinada concepción del *Big Data*. En primer lugar, el "Data" refiere a "algo dado" (en inglés,

something given). Llegada la década del 40, esta acepción fue complementada a partir del avance de la digitalización y utilizada para describir cualquier tipo de información que fuera utilizada y almacenada en el contexto informático. En contraste con el término "data", que apareció fundamentalmente vinculado al ámbito de la matemática y la teología, el *Big Data* encuentra sus primeros usos en la esfera de los negocios. La nueva infraestructura técnica y la capacidad de interrogar grandes volúmenes de datos con el fin de realizar predicciones, marcaron los 60: *"Big Data marked a suggested shift from relational database management systems to platforms that offered long-term performance advantages over traditional solutions"* (p. 1694). Esta idea de predicción sobrevuela aún en el discurso científico, en una suerte de anhelo utópico. Como sugiere Diviani (2018), este anhelo "se sustenta en uno de los aspectos básicos del dispositivo técnico que despierta gran fascinación: la supuesta facultad de pronosticar los fenómenos futuros. Sin embargo, ¿es posible la predicción de los fenómenos sociales y humanos?" (p. 14).

Distinto es el caso, por ejemplo, de aquellos estudios que, a partir de la utilización del *Big Data* y motorizados en tiempo real, marcan la evolución de un fenómeno; tal el caso actual de la propagación mundial del COVID-19[2] realizado por el Johns Hopkins Center for Systems Science and Engineering. El mapa fue compartido públicamente a mediados de enero de 2020 y se desarrolló con el fin de proveer a investigadores, periodistas, agencias de gobierno y ciudadanía en general, una fuente de datos confiables y actualizados en tiempo real a nivel mundial. Todos los datos se han puesto a disposición de manera gratuita a través de un repositorio de GitHub y se puede acceder a su código para ser compartido en distintas plataformas. El mapa es actualizado tanto de manera automática como manual y verificado con los informes oficiales que cada organismo de salud de los distintos países va dando a conocer.

El cambio no es para nada menor: fenómenos que, hasta el momento, parecían inabarcables, comienzan a manifestarse como, al menos, posibles de ser estudiados. Claro está que la utilización del *Big Data*, en tanto metodología de análisis, es necesaria y útil en tanto se encuentre en concordancia con las preguntas de investigación y los objetivos que guían determinado estudio. En todo caso, como plantea Diviani (2018):

2 https://coronavirus.jhu.edu/map.html (última consulta: 13 de marzo de 2020)

> Lo que habría que evaluar, en realidad, son dos ideas que acompañan las aseveraciones grandilocuentes y entusiastas, que están más próximas al marketing que al campo científico: la idea de que lo grande, the Big, es un valor positivo en sí mismo y que el dato es una porción de la `realidad´ neutra y objetiva (Diviani, 2018: 20-21).

Volveremos sobre la idea de neutralidad y objetividad en páginas siguientes. En este sentido, varios autores coinciden en caracterizar al *Big Data* a partir de "las 3 V": volumen, velocidad y variedad. Tascón (2013) sugiere, por su parte, una cuarta V: la visualización, en tanto una parte importante conectada al *Big Data* tiene que ver con la forma en la que podemos ver esos datos. Y Sosa Escudero (2019) suma una quinta: la veracidad, "término que se refiere a que la naturaleza ruidosa y espontánea de los datos del *Big Data* contrasta con la de los datos burocráticos o de encuestas tradicionales, usualmente sometidos a puntillosos ejercicios de validación" (p. 32). Ahora bien, ciñéndonos al ámbito específico de las ciencias sociales, podemos diferenciar –al menos– dos tipologías de datos: por un lado y siguiendo a Manovich (2012), el *surface data* y el *deep data*; y, por el otro, varios autores (Boyd y Crawford, 2012, Kitchin & Lauriault, 2015; Meneses Rocha, 2018) diferencian, también, el *Big Data* del *small data*.

En el primer caso, se trata de datos vinculados a distintas esferas de conocimiento. Los datos de superficie, se vinculan a metodologías cuantitativas que posibilitan el acceso a datos sobre muchos; al tiempo que los datos profundos están ligados a metodologías cualitativas que han permitido conocer problemáticas particulares pero con mayor nivel de especificidad. En cuanto a la segunda distinción, la captura del *small data* "ocurre de manera controlada y responde a un diseño estadístico y conceptual *ex profeso*" (Meneses Rocha, 2018: 422). Las "tres V" contienen amplias diferencias en ambos casos. En el *small data*, el volumen y la variedad es limitada y la velocidad, lenta. En el *Big Data*, por el contrario, el volumen es casi ilimitado, la velocidad es rápida y continua y la variedad es amplia.

Así como resaltamos la importancia que significa, para la ciencia en general y para las ciencias sociales en particular, la posibilidad de acceder a un gran número de datos, no podemos pasar por alto el valor económico y político que estas grandes cantidades de datos significan. Porque, claro está, los datos –a diferencia de los que se generan al responder una encuesta– son "anárquicos y espontáneos" (Sosa Escudero, 2019: 31). Encender el GPS del

celular para llegar a determinado lugar, sacar una foto que se encuentra geolocalizada, abrir una *app*, genera, indefectiblemente, un conjunto de datos que, originalmente, no fueron producidos más que con el propósito de llegar a hora a un lugar o tomar una foto de algo interesante. Ligado a esto, están los problemas y disputas en torno a la privacidad de los datos que conlleva. En su mayoría, los datos pertenecen a las propias empresas tecnológicas –la mayoría de ellas, situadas en Estados Unidos: Google, Facebook, Amazon, Apple y Microsoft– y, obtenerlos con fines de producción de conocimiento, se vuelve una tarea dificultosa, cuando no, imposible. En cuanto a la privacidad, el escándalo de Cambridge Analytica de 2018 dejó al desnudo las perversas maniobras de utilización de datos para fines políticos, sin consentimiento de los usuarios. Es menester recalcar, además, que el ingreso a cualquier red social requiere la aceptación de un conjunto de términos y condiciones que, sea por su longitud o por su descripción técnica –cuasi oscurantista–, muy pocas personas son capaces de comprender. En este mismo sentido, se genera una especie de ratonera: la aceptación de esos términos y condiciones se vuelven la condición necesaria para el acceso a las plataformas.

Habiendo realizado este somero recorrido, podemos definir al *Big Data*, tal como lo hacen Boyd y Crawford (2012) en tanto fenómeno cultural, tecnológico y académico que descansa en la interacción de:

> *Technology: maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets. (2) Analysis: drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims. (3) Mythology: the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy* (Boyd y Crawford, 2012: 663).

Un aspecto tecnológico, analítico y mitológico se desprende de la definición planteada *supra* que generaría, tal como lo expresan las autoras, un aura de verdad, objetividad y exactitud. Es esto, ante todo, lo que pareciera reavivar la discusión entre los abordajes cuantitativos y cualitativos; o lo que podríamos llamar, siguiendo la literatura especializada en el tema, una nueva forma de empirismo: una preeminencia de los datos por sobre la teoría. Coincidimos, como plantea Tagnin (2019) en que los interrogantes

que se abren en las ciencias sociales a partir de la llegada del *Big Data* "reviven históricos debates sobre los criterios para fundamentar los dominios de objetos legítimos en el discurso científico" (s/p). La posibilidad efectiva de contar con grandes cantidades de datos, que replican relaciones en distintos niveles, podría verse como una nueva etapa en la que la apelación a estos permite encontrar *evidencias* de la vida social[3].

Detengámonos pues, en los sentidos asociados al *Big Data*. Como hemos dicho, los datos parecieran contar, *per se*, con una gran cuota de verdad, objetividad y neutralidad. En principio es necesario tener en cuenta que todo dato no significa una porción extraída, sin mediaciones, de la realidad. Por el contrario, los datos que son utilizados para realizar investigaciones científicas, son construcciones sociales, representaciones de determinados fenómenos que "no existen fuera de las ideas, los instrumentos, las práctica y el contexto que enmarca su creación e interpretación" (Meneses Rocha, 2018: 424). La mirada, crítica y reflexiva, de parte de los investigadores continúa siendo tan necesaria como los datos que nutren los trabajos. Además, otro punto importante tiene que ver con la representatividad de los datos extraídos, en función del fenómeno a analizar, las preguntas de investigación y los objetivos de cada estudio. Como decíamos en un trabajo anterior (Gindin y Busso, 2018), la alerta respecto de esta representatividad tiene que ver "no sólo por lo potencialmente inmanejable de los grandes volúmenes de información en base a métodos de análisis cualitativos, sino también porque se trata de un universo de datos en permanente expansión y del cual no conocemos las fronteras" (p. 32). De aquí la oportuna pregunta que se hacen Cingolani y Fernández (2018): "¿con qué criterio construir representatividad si no sabemos dónde toca fondo el universo?" (p. 160). Por ejemplo, para el caso de Twitter, Gaffney y Puschmann (2014) alertan sobre la sub o sobre valoración que se le puede otorgar a los datos extraídos de cuentas de Twitter en relación a la población en general. Esto es, ¿cuán representativos son los usuarios de Twitter en función del conjunto de la sociedad? Y, por otro lado, si construimos hipótesis en relación a la población de usuarios de Twitter, ¿cómo medimos aquellos que sólo postean eventualmente, de aquellos que lo hacen todo el tiempo?

3 Puede tomarse como ejemplo la sociología y la utilización de las encuestas con el fin de poder explicar ciertas conductas sociales.

En este marco de sentidos asociados al *Big Data*, Puschamn y Burgues (2014a), analizan dos metáforas utilizadas en la prensa[4] que nos permiten pensar la proyección de un conjunto de ideas que circulan socialmente. La primera de ellas es la que hace pensar al *Big Data* como una *fuerza natural que debe ser controlada*: una masa uniforme, donde conviven datos esenciales, valiosos, difíciles de controlar y ubicuos. La vinculación con el agua supone, además, una adjudicación de valor neutral: el agua es, *per se*, insípida e incolora. Ahora bien, la posibilidad de ahogarse en ese torrente de datos es también un peligro inminente y constante. La segunda de las metáforas analizadas es la del *Big Data* como *alimento o combustible que debe ser consumido* y aquí los autores señalan dos posibles desprendimientos. De un lado, la idea de que el *Big Data* es un alimento indispensable para la supervivencia; del otro –y como contracara de la misma moneda– el *Big Data* se proyecta como un combustible que alimenta, por ejemplo, a las empresas. Como afirman los autores, *"both food and fuel must be consumed to exist and to move forward rather than being consciously used"* (p. 1700).

El arribo y consolidación del *Big Data* ha generado, además, la configuración de nuevos campos disciplinarios tales como las humanidades digitales, las ciencias sociales computacionales y las ciencias de los datos[5], entre otras, en una suerte de desdibujamiento de las fronteras disciplinares. Por tanto, rápidamente se advierte una necesidad de imbricación entre abordajes "artesanales" y saberes vinculados a las ciencias computacionales; imbricación que, además, implica, de parte de los cientistas sociales, "comprender de forma realista y crítica la transformación de datos en conocimiento útil para la sociedad" (Meneses Rocha, 2018: 416). Los métodos de aprendizaje automático, que se sitúan en la frontera entre la estadística y la computación, parecieran resolver –al menos por ahora– el manejo de estos volúmenes de datos: se supone que un proceso computacional aprende, sobre la base de algún criterio previamente fijado, de manera automática. Con

4 Los ejemplos son tomados de *The Wall Street Journal*, *Forbes*, and *Business Insider*, *Wired and Computerworld*, *The Chronicle Herald*, *USA Today*, *World Future Society*, y *Booz & Company*.

5 Para el caso específico de Twitter, sugerimos la lectura de Gaffney y Puschmann (2014b) en el que realizan un recorrido por las diferentes herramientas actualmente disponibles para recolectar datos, marcando de este modo sus potencialidades y sus limitaciones. Para conocer otras técnicas utilizadas no necesariamente vinculadas a Twitter, recomendamos Arcila-Calderón, Barbosa-Caro y Cabezuelo-Lorenzo (2016).

cada vez más fuerza, el abordaje interdisciplinario se vuelve un punto de pasaje obligado para la realización de trabajos serios que no descuiden las limitaciones que hemos mencionado a lo largo de estas páginas.

3. Memoria y archivo en la era del *Big Data*

La presencia del *Big Data*, tanto como insumo de trabajo como dilema teórico-metodológico, enfrenta a los analistas sociales a las problemáticas que hemos mencionado más arriba, pero no sólo. Además de los interrogantes ligados a la construcción de los corpus de análisis y a la *expertise* necesaria para gestionarlos, la existencia de grandes volúmenes de información sobre los gustos, los consumos, las opiniones, los desplazamientos, de grandes cantidades de personas, imponen una reflexión obligada sobre la *temporalidad* de esos contenidos; o más bien, sobre cómo éstos conforman un *archivo*, en el sentido de registro de la actividad de los usuarios en Internet.

Como ya hemos mencionado en otro lugar (Gindin y Busso, 2018), los cientistas sociales que deciden encarar investigaciones empleando *Big Data* se ven obligados a resolver espinosas cuestiones sobre el acceso a éste, su relevamiento y sistematización. Dichas operaciones imponen además un reconocimiento de que se trata de una particular dimensión temporal que se actualiza: no solamente por tratarse de contenidos producidos inevitablemente en un pasado, más o menos reciente o más o menos lejano, sino porque el acceso a éstos implica ciertas omisiones, ciertas lagunas, ciertos *olvidos*.

No se trata en este caso de una desmemoria selectiva, sino de una imposibilidad técnica: muchas veces, y aunque exista una especie de convencimiento extendido de que es posible acceder a *todos* los materiales presentes en Internet, es en realidad imposible recuperar la totalidad del material producido. Es ejemplificador al respecto el caso de Twitter, donde la propia plataforma, si bien habilita a recolectar los tuits publicados en ella, permite acceder sólo a una parte de éstos[6].

6 Como señalan Gaffney y Puschmann (2014) una de las grandes dificultades de los estudios basados en datos provenientes de Twitter tiene que ver con la propia infraestructura técnica de la plataforma. Actualmente, no hay manera de saber cuán completo es el recupero de un conjunto de datos en función de lo que fue publicado originalmente: *"without firehose access, researchers rely entirely on Twitter to provide a representative sample of what is there"* (p. 65)

Por supuesto, el material recopilado –incluso con sus omisiones– sigue siendo cuantioso, incluso inabarcable. Pero reconocer esas discontinuidades nos permiten poner en relieve las complejidades de un rol que, quizás involuntaria aunque obligadamente, toca asumir a quienes empleen este tipo de materiales: el de una especie de nuevo archivista, un corresponsable de la gestión de la *memoria* de los espacios digitales.

Entendemos al archivo como una construcción realizada por el investigador, con sus particulares lógicas de estructuración y sus mecanismos de resguardo: ambos delimitan la exigencia de un orden de lectura preciso, en estrecha vinculación con el espacio y la época en los que se lo hace intervenir. Así, el archivo "no es ni reflejo del acontecimiento ni tampoco su demostración o prueba. Siempre debe ser trabajado mediante cortes y montajes incesantes con otros archivos" (Didi-Huberman, 2007: 7). Lo que se pone en cuestión con estas afirmaciones, entonces, es que la presencia de grandes cantidades de datos, muchas veces capaces de brindar información hiperdetallada y minuciosa sobre las acciones y las conductas, no implican necesariamente un reservorio exhaustivo de todo aquello que sucede.

Por ello, debemos reconocer que –en realidad– aquello que se conserva y se resguarda es producto de distintas operaciones: algunas más automáticas, de recuperación de datos, y otras más bien intencionales, como la decisión de qué recuperar y, luego, de qué conservar. En una sociedad crecientemente mediatizada (Verón, 2013), donde el ritmo de la vida social está directamente relacionado con la presencia de los medios –y, agregamos, con las tecnologías a ellos ligadas–, el reconocimiento de esa actividad archivística no es banal, ya que implica hacer inteligibles (Freire, 2009) esas grandes cantidades de información preservada, donde aparecen también múltiples voces ínfimas y cotidianas que se vuelcan a distintos espacios de Internet, y pueden obtener así una precaria garantía de resguardo... o al menos eso suponen.

Precisamente, en las sociedades mediatizadas tal como las conocíamos hasta ahora han sido principalmente los medios masivos los que han sido señalados como los vehículos de la memoria de la experiencia humana, proponiendo rememoraciones e interpretaciones de un pasado común, en el que participa la memoria individual (Lavabre, 2007). Se trata de una *cultura de la memoria* (Huyssen, 2002), con su contracara de un enorme temor al olvido y a la amnesia (Guasch, 2005). En la actualidad, sin embargo, podemos aventurar que es el *Big Data* aquél que es señalado como el depositario de esa memoria, aunque confundiendo aquí el mero almacenamiento de

información con las actividades de accesibilidad y síntesis que ella demanda, e invisibilizando las omisiones y discontinuidades en la información de las que dimos cuenta más arriba.

Con estos resguardos, podemos afirmar que efectivamente el *Big Data* es un potente ámbito de archivo de la memoria contemporánea, que nos permite conservar una cantidad y minuciosidad de informaciones hasta hace poco tiempo impensables. Sin embargo, este reconocimiento no implica obturar otros espacios de resguardo en los que aparecen las voces de quienes no participan del intercambio digital; cada vez un número menor, es cierto, aunque todavía consistente: si vamos por ejemplo al caso de los datos recuperados de Internet, encontramos que para enero de 2020 el 59% de la población global estaba conectada[7]. Junto a ellos, los matices, los reenvíos contextuales y de sentido, y también ciertos diálogos que involucran diversas materialidades significantes, que no pueden ser recuperados en esa gran memoria de datos.

Por añadidura, y como anticipábamos párrafos arriba, la presencia de esta gran cantidad de datos y su almacenamiento no implica que se cuente con las adecuadas políticas de resguardo y de inteligibilidad. En primer lugar, lo cuantioso de la información almacenada requiere de grandes espacios físicos (los así llamados *data center*) donde ubicar los servidores y los discos duros donde se conservan esos datos, en un *in crescendo* que requiere cada vez de nuevas capacidades materiales[8]. En segundo término, y recuperando aquí el concepto de *arqueología* de Foucault, podemos entender que en torno al *Big Data* existen prácticas específicas que le dan un orden, una organización, una distribución interna: "repartiéndolo en niveles, estableciendo series, distinguiendo lo que es pertinente de lo que no lo es, señalando elementos, definiendo unidades, describiendo relaciones y elaborando discursos" (Guasch, 2005: 160). Aquello que del archivo se hace visible, aprehendible, es entonces el resultado de operaciones que, en términos de la investigación social, se vinculan con la definición de un objeto de estudio, las hipótesis de trabajo en

7 Fuente: Digital 2020. Informe anual de We Are Social, disponible en https://wearesocial.com/es/blog/2020/01/digital-2020-el-uso-de-las-redes-sociales-abarca-casi-la-mitad-de-la-poblacion-mundial (última consulta: 20 de marzo de 2020)

8 Ver, por ejemplo, https://www.ibm.com/blogs/systems/donde-esta-el-almacenamiento-de-datos-con-big-data/ y https://fractaliasystems.com/2016/08/03/big-data-donde-se-almacena/ (última consulta: 20 de marzo de 2020)

relación a éstas, y las técnicas de sistematización y análisis puestas en juego. Es decir: es necesaria una operación voluntaria y hasta sistemática para *hacer decir* a esa gran cantidad de información disponible.

Probablemente, y a este punto del recorrido, aquello que sí podemos afirmar es que con el *Big Data* estamos de frente a un nuevo cambio en las formas de reconocimiento e interpretación del mundo. En él, creemos, se delinea no sólo la construcción de experiencias sensibles relativas a ese estar con otros, sino que también se define qué temas son dignos de atención, convirtiéndose en un instrumento privilegiado de saber sobre ese estar en común. Al fin de cuentas, podría decirse parafraseando a Sorlin (2004), que estos dilemas darán forma al debate sobre qué caracteriza a la sociedad moderna, preservando o bien dejando afuera *datos* sobre ella, a los que accederán las generaciones futuras.

4. Conclusiones

Denominar al tramo final de este trabajo como "conclusiones", en base a lo expuesto previamente, no deja de ser un contrasentido. Los escenarios posibles que se abren en torno al empleo del *Big Data* son cuanto menos amplios, y se ramifican en diversos ámbitos. Por quedarnos en el de las ciencias sociales, la economía, la política, la sociología y las ciencias de la comunicación, entre otras, probablemente estén asistiendo a un cambio profundo e irreversible en lo que respecta a la misma definición de sus objetos de estudio y las estrategias metodológicas diseñadas a tal fin.

Sin embargo, y recapitulando lo expresado en estas páginas, sabemos que el *Big Data* (y sus correlativos *small data*, *deep data*, etc.) no es un mero agregado neutral de información disponible para todos, sino que se trata de un complejo fenómeno cultural, tecnológico y académico, que demanda explícitos posicionamientos epistemológicos de parte del investigador y no sólo. Si bien es cierto que muchos estados nacionales y muchos gobiernos los consideran cada vez más un insumo vital a la hora de pensar sus políticas públicas, y que incluso mecanismos de representación política –para el caso– parecen estar siendo modificados por la posibilidad de recolectar y emplear grandes cantidades de datos relativos a las conductas más ínfimas de los ciudadanos, también es necesaria aquí una reflexión alerta.

En ese sentido, es notorio el entusiasmo por aquello que se suele entender como la fabulosa oportunidad de conocer las conductas y valores de los ciudadanos, a partir de las huellas que éstos dejan en los distintos espacios digitales en los que participan. Además, las trazas que los usuarios de medios sociales, por ejemplo, van dejando en el camino, se vuelven rápidamente un objeto de consumo; se compran y venden datos dando forma a un ecosistema comercial (Puschmann y Burguess, 2014b). Esta posibilidad –cierta– no debe invisibilizar cuánto de opaco hay en ella: como mencionábamos previamente, los datos no hablan *per se* ni son ajenos a los intereses de la investigación o de la política; y yendo más allá, tampoco son capaces de dar cuenta de la totalidad del mundo social.

Aquello que queda afuera, ya sea por las elipsis temporales propias de la actividad de recolección, o bien por tratarse de actividades sin huella en el mundo digital, también debería ser entendido como parte de la discursividad de una época, aunque no aparezca considerado a la hora de hablar de macrodatos y afines. Las metáforas con las que se alude al *Big Data* (aquella que lo toma como una fuerza de la naturaleza, y la que lo compara con un combustible a ser consumido) nos permiten pensar los matices de ese concepto, aunque no pueden abarcar, justamente, lo que queda por fuera de ellas.

Pensar la tarea de los científicos sociales en este contexto implica no eludir esas disputas. Probablemente sea exagerado plantear que estamos frente a un rotundo cambio de época sólo en base a la presencia del *Big Data*, pero sí podemos afirmar que sus implicancias son un punto de no retorno en los alcances que se pretenden para la ciencia.... y para la tarea política a ella vinculada. Ejemplos abundan, como ser, la posibilidad de prever el recorrido de un virus (tal como se ha buscado hacer frente a la pandemia del COVID-19), en base a los macrodatos recolectados sobre los desplazamientos y los contactos de las personas; o la propuesta de reconocer el (mal) humor social frente a un determinado argumento de interés público. Sin embargo, no hay que olvidar que no *todos* pueden realizar tales previsiones, ni que *todo* esté contemplado en ellas. Nuevamente, la disponibilidad de tales datos no es la única condición requerida: también lo es la destreza cognitiva y técnica para aplicarlos, y –por qué no– la disponibilidad económica y material para hacerlo.

Sin dudas, no solamente el *Big Data* hablará de nosotros en un futuro sino también estas discusiones, como ecos de un escenario en incierta transformación.

5. Bibliografía

ARCILA-CALDERÓN, C., BARBOSA-CARO, E. y CABEZUELO-LORENZO, F. (2016). "Técnicas Big Data: análisis de textos a gran escala para la investigación científica y periodística". en *Revista El Profesional de la Información. International Journal of Information and Communication*, España: Ediciones Profesionales de la Información S.L., pp. 623-631. Disponible en http://www.elprofesionaldelainformacion.com/contenidos/2016/jul/12.pdf

BEDOYA, M. E., WAPPENSTEIN, S. (2011). "(Re)pensar el archivo. Presentación del dossier". En *Íconos. Revista de Ciencias Sociales*. Ecuador: FLACSO Ecuador, vol. 41, pp. 11-16. Disponible en https://revistas.flacsoandes.edu.ec/iconos/article/view/386

boyd, D. y CRAWFORD, K. (2012). "Critical questions for Big Data". En *Information, Communication & Society*, London: Routledge, vol. 15, n° 5., pp. 662-679.

DIVIANI, R. (2018). "Consideraciones epistemológicas, teóricas y críticas en relación al Big Data". En BISELLI, R. y MAESTRI, M. (eds.) (2018). *La mediatización contemporánea y el desafío del Big Dat.*: Rosario: UNR Editora, pp. 11-34. Disponible en https://cim.unr.edu.ar/publicaciones/1/libros/137/la-mediatizacion-contemporanea-y-el-desafio-del-big-data

CINGOLANI, G. y FERNÁNDEZ, M. (2018). "Objeto, objetos, corpus, sistemas. Lo que se deshace y lo que hacemos cuando analizamos discursividad en la crisis de mediatización contemporánea". En BISELLI, R. y MESTRI, M. (eds.) (2018). *La mediatización contemporánea y el desafío del Big Data*. Rosario: UNR Editora, pp.148-176. Disponible en https://cim.unr.edu.ar/publicaciones/1/libros/137/la-mediatizacion-contemporanea-y-el-desafio-del-big-data

FREIRE, C. (2009). "Artistas/curadores/archivistas: políticas de archivo y la construcción de las memorias del arte contemporáneo". En FREIRE, C. y LONGONI, A. (orgs.). *Conceptualismos del Sur/Sul*. São Paulo: Annablume -USPMAC- AECID, pp. 64-82.

GAFFNEY, D. y PUSCHMANN. C. (2014). "Data Collection on Twitter". En WELLER, K., BRUNS, A., BURGUESS, J. MAHRT, M. & PUSCHAMNN, C. (eds.) (2014). *Twitter and society*, New York: Peter Lang Publishing, pp. 55-68.

DIDI-HUBERMAN, G. (2007). "El archivo arde", traducción realizada por la Cátedra de Filología Hispánica de la Universidad Nacional de La Plata.

Disponible en https://filologiaunlp.files.wordpress.com/2012/05/el-archivo-arde1.pdf

GINDIN, I. y BUSSO, M. (2018). "Investigaciones en comunicación en tiempos de Big Data: sobre metodologías de análisis y temporalidades en el abordaje de redes sociales". En *Revista adComunica: Revista Científica de Estrategias, Tendencias e Innovación en Comunicación*. Castellón: Asociación para el Desarrollo de la Comunicación adComunica y Universitat Jaume I, nº15, pp. 25-43. Disponible en http://www.adcomunicarevista.com/ojs/index.php/adcomunica/article/view/453

GUASCH, A.M. (2005). "Los lugares de la memoria: el arte de archivar y recordar". En revista *Materia*: España, Universidad de Barcelona, n° 5, pp. 157-183. Disponible en http://revistes.ub.edu/index.php/materia/article/view/11382

HUYSSEN, A. (2002). "Pretéritos presentes: medios, política y amnesia": *En En busca del futuro perdido. Cultura y memoria en tiempos de globalización*. México: Fondo de Cultura Económica, pp. 13-40.

JELIN, E. (2002). *Los trabajos de la memoria*. Buenos Aires: Siglo XXI.

KITCHIN, R. & LAURIAULT, T. (2015). "Small data in the era of Big Data". En *GeoJournal Spatially Integrated Social Sciences and Humanities*. Países Bajos: Kluwer Academic Publishers, nº 80, pp. 463–475.

LAVABRE, M.C. (2007). "Maurice Halbwachs y la sociología de la memoria". En PEROTIN-DUMON, A. (dir.), *Historizar el pasado vivo en América Latina*. Disponible en http://www.historizarelpasadovivo.cl/downloads/lavabre.pdf

MANOVICH, L. (2012). "Trending: The Promises and the Challenges of Big Social Data". En GOLD, M. (2012). *Debates in the Digital Humanities*. Minnesota: University of Minnesota Press.

MENESES ROCHA, M.E. (2018). "Grandes datos, grandes desafíos para las ciencias sociales". en *Revista Mexicana de Sociología*. Ciudad de México: Universidad Nacional Autónoma de México-Instituto de Investigaciones Sociales, nº2 (abril-junio, 2018), pp. 415-444. Disponible en http://revistamexicanadesociologia.unam.mx/index.php/rms/article/view/57723/51185

PUSCHMANN, C. y BURGUESS, J. (2014a). "Metaphors of Big Data". En *International Journal of Communication*. Califormia: University of Southern California, vol, 8(4), pp. 1690-1709. Disponible en https://ijoc.org/index.php/ijoc/article/view/2169

_____ (2014b). "The politics of Twitter Data". En WELLER, K., BRUNS, A., BURGUESS, J. MAHRT, M. & PUSCHAMNN, C. (eds.) (2014). *Twitter and society*. New York: Peter Lang Publishing, pp. 43-54.

SORLIN, P. (2004). *El 'siglo' de la imagen analógica. Los hijos de Nadar.* Buenos Aires: La Marca Editora.

SOSA ESCUDERO, W. (2019). *Big Data. Breve manual para conocer la ciencia de los datos que ya invadió nuestras vidas*, Buenos Aires: Siglo XXI.

TAGNIN, J. (2019). "Big Data y ciencias sociales"- en *Revista Bordes*. Disponible en: https://revistabordes.unpaz.edu.ar/big-data-y-ciencias-sociales/

TASCÓN, M. (2013). "Introducción. Big Data. Pasado, presente y futuro". En *Revista Telos. Cuadernos de Comunicación e Innovación*. España: Fundación Telefónica, nº 95, pp. 47-50. Disponible en https://telos.fundaciontelefonica.com/archivo/numero095/pasado-presente-y-futuro/

VERÓN, E. (2013). *La semiosis social 2. Ideas, momentos, interpretantes.* Buenos Aires: Paidós.