



Hacia un análisis de los marcos epistémicos del big data

Towards an analysis of the epistemic frameworks of big data

Gastón Becerra (gastonbecerra@sociales.uba.ar) Consejo Nacional de Investigaciones Científicas y Técnicas (Buenos Aires, Argentina) <https://orcid.org/0000-0001-9432-8848>

José Antonio Castorina (ctono@fibertel.com.ar) Consejo Nacional de Investigaciones Científicas y Técnicas (Buenos Aires, Argentina) <http://orcid.org/0000-0003-1724-9315>

Abstract

In this paper we explore some epistemological discussions about big data. As our main epistemological concept, we use the notion of Epistemic Framework formulated by Piaget and García, and reviewed by us in more recent works, which seeks to problematize the conditioning of ethical and political values, as well as other regulations, in the development of scientific knowledge. We propose to order the debates in the recent literature on big data through various dualities, such as data/interpretation, automatic/non-automatic, or causality/correlation, around which it is possible to view two specific epistemic frameworks, tending to split these terms, or to their integration. On this last mode of focusing the debates, we identify some epistemological challenges of research with big data, especially in the computational social sciences.

Key words: big data, epistemic framework, constructivism, computation.

Resumen

En este trabajo queremos indagar sobre algunas cuestiones epistemológicas que se discuten en torno al fenómeno del big data. Como principal herramienta de análisis epistemológico utilizamos a la noción de Marco Epistémico formulada por Piaget y García, y revisada por nosotros en trabajos más recientes, para problematizar los condicionamientos de los valores éticos y políticos, así como de otras regulaciones, en el desarrollo del conocimiento científico. Para ello, proponemos ordenar los debates que se registran en la literatura reciente sobre big data a través de diversas dualidades, tales como dato/interpretación, automático/no-automático o causalidad/correlación, en torno a las cuales vamos a argumentar que es posible pensar dos marcos epistémicos específicos tendientes a la escisión de estos términos o a su integración. Abogando por esta última manera de focalizar los debates, hacia el final del trabajo identificamos algunos desafíos epistemológicos de la investigación con big data, especialmente en las ciencias sociales computacionales.

Palabras clave: big data, marco epistémico, constructivismo, computación.



Introducción

Se suele entender por big data la disponibilidad de grandes volúmenes de datos, en formatos diversos, pero mayormente con soporte digital, que pueden ser analizados y explotados para transformar (presumiblemente) todos los aspectos de nuestra realidad social, material o subjetiva. Este artículo hace foco en lo que se denomina la mitología del big data, según la cual “grandes conjuntos de datos ofrecen una forma superior de inteligencia y conocimiento hasta ahora imposible, con un aura de verdad, objetividad y precisión” (boyd y Crawford 2012:663). Esta es una creencia de naturaleza epistémica que supone una concepción de lo que es el conocimiento, cómo se construye, a qué objetivos sirve, y con qué criterios y valores lo evaluamos socialmente. Desde esta mitología se sostiene el mandato de que todos los aspectos de la realidad pueden y deben convertirse en datos; que el conocimiento se construye a partir de patrones y correlaciones, llegando a ser innecesaria la construcción de teorías, interpretaciones y explicaciones; que mayor cantidad, diversidad y complejidad de los datos permite un conocimiento menos propenso a errores y sesgos; que un mejor conocimiento es el más accionable y el que habilita una manipulación más efectiva; que el conocimiento del futuro es posible; o que supone una revolución científica o un nuevo paradigma que supera la manera tradicional de hacer ciencia. Iniciar una discusión epistemológica de estas creencias es difícil, porque el discurso en torno al big data es difuso y polifónico y no se puede atribuir a un portavoz claro para plantear una controversia. Acerca del big data y sus capacidades se han pronunciado actores muy diversos, que se relacionan con el conocimiento de manera heterogénea, y cuyas comunicaciones se dan en canales que raras veces se integran. Entre estos podemos mencionar a científicos y filósofos de ramas del conocimiento muy diversas, así como a líderes políticos y CEOs de empresas, gurús tecnológicos que escriben *best sellers* y expertos y vendedores de soluciones cuyos discursos promocionales se reproducen en medios masivos.

En un intento por ordenar la discusión epistemológica en torno al big data, en este trabajo proponemos como herramienta de análisis a la noción de marco epistémico (en adelante ME) del programa epistemológico del constructivismo piagetiano, que busca problematizar diversas dimensiones en la relación entre conocimiento y sociedad. Basándonos en la manera en que este enfoque ha servido para reflexionar acerca de las condiciones de desarrollo de algunas ramas de la psicología, aquí buscamos explorar una clave interpretativa que ordene y refocalice las controversias epistemológicas en torno al big data y posibilite una mirada más compleja del fenómeno para identificar sus principales desafíos, especialmente para las ciencias sociales.

Nuestro trabajo se organiza de la siguiente forma. En el primer apartado nos referimos brevemente al entramado social en el que se origina el discurso epistémico del big data y su vinculación con el conocimiento científico. En el segundo, introducimos la noción de ME del constructivismo, haciendo foco en cómo se lo ha utilizado como herramienta de análisis epistemológico de desarrollos científicos contemporáneos. En el tercero, reseñamos algunas dicotomías que podemos encontrar en la literatura epistemológica reciente sobre el big data, tales como dato/interpretación, automático/no-automático, o causalidad/correlación, en torno a las cuales, propondremos, operan dos ME con concepciones del mundo opuestas: una escisionista (ME), que se caracteriza por oponer los términos de las dualidades o reducir unos a otros y otra relacional (MER) que entiende que los términos se encuentran integrados y que una perspectiva contraria obstaculiza el planteo de los problemas significativos. En el cuarto, y desde esta última posición, identificamos algunos desafíos epistemológicos de la investigación con big data, especialmente en las ciencias sociales



computacionales. Finalmente, en las conclusiones evaluamos las contribuciones y las limitaciones del uso de la noción de ME para pensar el big data y esbozamos algunas preguntas pendientes para futuras discusiones.

1. Big data: un discurso social y su impacto en la ciencia

Las primeras menciones del big data lo ubican en el campo de la tecnología hacia finales de la década de los 90, refiriéndose a los desafíos computacionales del manejo de grandes datos, como la necesidad de desarrollar herramientas e infraestructuras informáticas. A falta de una definición más precisa, se lo asoció a la fórmula de las “3 Vs del big data” para indicar que involucra grandes *volúmenes* de datos, *variedad* de formatos de datos y la necesidad de un procesamiento *veloz*. Hacia el año 2000, el big data empieza a ser tematizado en el mundo de la consultoría y a estos sentidos se le suman otros que, en lugar de remitir a características o propiedades ontológicas y epistemológicas de los *datasets*, pusieron el foco en sus posibles usos. Por caso, desde el mundo del *management* se comenzó a hablar de *valor* para referir a la posibilidad de explotar los registros de producción y transacciones comerciales como insumos en el diseño de productos y estrategias que generen una ventaja competitiva. Para la década siguiente, el big data ya es reconocido como un fenómeno de alcance ubicuo al que se dedican notas en medios masivos y libros de divulgación para introducirlo al público general. Se habla de una “era del big data” y de una “revolución que transformará cómo vivimos, trabajamos y pensamos”.

El proceso social subyacente es la “datificación”, es decir, el registro y la conversión en datos cuantitativos de diversos aspectos de nuestra vida para su posterior explotación que, si bien puede rastrearse hasta los albores de la modernidad, ha obtenido un enorme impulso en los últimos años gracias a la participación (muchas veces forzada o coaccionada) en plataformas digitales que registran nuestras interacciones, relaciones, consumos, intereses y gustos. De acuerdo con Zuboff ([The age of surveillance capitalism](#)), sobre este sustrato es erigido un nuevo orden económico (un capitalismo de vigilancia), cuyo *commodity* es la información personal y la experiencia privada de cada uno de nosotros, la cual se comercia en un mercado de comportamientos futuros. En este régimen, los individuos no son vendedores ni compradores, sino materia prima sujeta a una lógica de extracción y especulación, en la que los mismos dispositivos que nos registran buscan coaccionar y automatizar nuestros comportamientos.

Los discursos más promocionales del big data lo han enfrentado, en algunos aspectos, a la ciencia tal como la conocemos hasta hoy. La formulación más provocadora de esta posición fue expresada por Chris Anderson en su artículo [The end of theory](#), en el que relata cómo Google logró dominar el mercado de la publicidad solo correlacionando datos, sin pretender probar hipótesis o generar explicaciones sobre por qué la gente hace lo que hace. Desde aquí se desprende un reclamo por un nuevo método para las ciencias en general (aunque los casos que reseña se limitan a la biología y la física). Reclamos similares se pueden encontrar en [Fourth paradigm](#), un volumen colectivo editado por Microsoft Research, basado en la propuesta de Gray de una síntesis entre ciencia y tecnología de la información, que no solo propone revisar el método científico, sino también repensar la organización y la comunicación de la ciencia, adoptando infraestructura y herramientas generadas por corporaciones tecnológicas.

En los últimos años, estos reclamos se han visto fuertemente cuestionados por parte de diversos filósofos de la ciencia y epistemólogos de distintos campos del conocimiento, que coinciden en



señalar que, por un lado, en materia de argumentos epistemológicos, no habría sustento para hablar de una revolución científica, sino que, más bien, se trata de un *revival* positivista que cuenta con grandes problemas para poder imponerse, tales como la dificultad de distinguir entre correlaciones espurias y significativas únicamente a partir de métodos inductivos. Aunque, por otro lado, también señalan que el big data es una gran oportunidad para renovar fuentes, técnicas y métodos de análisis, favoreciendo el desarrollo computacional de las distintas ramas de las ciencias, en la medida en que dicho esfuerzo se haga de una manera reflexiva y epistemológicamente argumentada. Por esto, para contribuir con esta empresa, en las siguientes secciones buscamos ordenar las discusiones epistemológicas desde herramientas conceptuales del constructivismo.

2. Acerca de la noción de marco epistémico

Para problematizar el big data apelamos a la noción de ME del constructivismo piagetiano revisado. Esta noción fue introducida hacia los 80 por Jean Piaget y Rolando García ([Psicogénesis e historia de la ciencia](#)) para referirse al componente sociocultural de una teoría del conocimiento (científico) constructivista. Sin embargo, la formulación más clara y difundida se encuentra en la reelaboración que Rolando García hizo de la epistemología piagetiana, cuando define al ME como “un sistema de pensamiento, rara vez explicitado, que permea las concepciones de la época en una cultura dada y condiciona el tipo de teorizaciones que van surgiendo en diversos campos del conocimiento” (García 2000:157). Este sistema de pensamiento sería equiparable con una cosmovisión del mundo o una concepción de la naturaleza y de lo social que forma parte del sustento ideológico de una época. También, a partir de la revisión de García, el ME se empieza a utilizar en otros campos del conocimiento, más allá del dominio de la historia de la ciencia para que la fue originalmente pensada. Por ejemplo, el mismo García ([Sistemas complejos](#)) la ha señalado como un componente central en la práctica de la investigación interdisciplinaria para el diagnóstico y la intervención en problemáticas referidas a sistemas complejos, vinculándola a los valores políticos que participan de la construcción conjunta del objeto de la investigación; o más recientemente, en el campo del análisis psicogenético se la ha utilizado para referir a los significados que envuelven a los objetos de conocimiento y que pueden suministrar una orientación o constituir un límite a las acciones cognoscitivas del individuo (Barreiro. [La creencia en la justicia immanente piagetiana](#)).

Otro de los usos que se le han dado a la noción de ME es en tanto herramienta conceptual para pensar problemas epistemológicos de la ciencia contemporánea. Aquí, se lo vincula con los supuestos ontológicos y epistemológicos, y especialmente con los valores no epistémicos que, indirectamente y sin determinarlos, condicionan la dirección de un programa de investigaciones o de un campo disciplinar. Este es, tal vez, el sentido más discutido de la noción. Por un lado, se lo ha utilizado especialmente para pensar problemas internos de la epistemología de las ciencias sociales (Yocelevszky. [Sociogénesis y sociología](#), Cortés. [Nociones de la epistemología genética aplicadas a temas de discusión en las ciencias sociales](#)) y uno de nosotros lo ha trabajado profusamente en relación con la psicología (Castorina. [Las concepciones del mundo y los valores en la investigación psicológica](#); [Los supuestos filosóficos en psicología del desarrollo](#)). Por otro lado, también ha merecido varios trabajos que buscaron ponerlo en diálogo con otros conceptos epistemológicos centrales, como paradigma de Kuhn, cosmovisión de Needham, marco normativo de Gómez, la reflexividad de Morin, entre otras (Becerra y Castorina. [El concepto de marco epistémico](#); [Acerca de la noción de “marco epistémico” del constructivismo](#); Tsou. [Genetic epistemology and Piaget's philosophy of science](#); García. [Similitudes y diferencias...](#); Rodríguez-Zoya. [Contribución a la crítica de la teoría de los sistemas complejos](#); Peruzzo. [La investigación científica](#); García-Bravo. [Rolando](#)



[García; Elementos para la construcción...](#)). Al recurrir al ME para pensar el big data, en este trabajo buscamos vincular las reflexiones epistemológicas de este nuevo campo del conocimiento con los desarrollados en estos antecedentes.

En todos estos casos, el ME sirve para problematizar las condiciones de contorno (de naturaleza social) en el desarrollo del conocimiento, de una manera no determinista, ya que sigue sosteniendo la primacía de la acción cognoscitiva del sujeto epistémico, a la vez que parece evitar los extremos de la dicotomía internalismo / externalismo. Así, por ejemplo, se puede asumir que un ME incide en la formulación de preguntas y problemas de la investigación, en la construcción de las unidades de análisis, la evaluación de los modelos explicativos que se ponen en juego, la selección y el diseño de métodos y técnicas empíricas, el alcance del uso o la aplicación de las teorías, y hasta las relaciones entre campos disciplinarios. Finalmente, un análisis que emplee la categoría de ME deberá poner el foco en que ciertos criterios y valores de la investigación (e.g. evaluaciones metodológicas) dependen de condiciones institucionales que se imponen a las prácticas de la investigación, y adquieren fuerza normativa a través de regulaciones políticas y administrativas, como condiciones para hacer una “verdadera ciencia” o una “ciencia rigurosa”.

Explicitar un ME y problematizar su intervención en la investigación es una tarea de los propios investigadores de un campo particular, resultante de la reflexión sobre su historia y situación, sus desarrollos teóricos y empíricos y de las condiciones en las que construyen conocimiento. Sin embargo, en las próximas secciones propondremos que en torno al big data se pueden observar algunas perspectivas epistemológicas que guardan semejanzas con los MEs que operan en algunas ramas de la psicología.

De acuerdo con los análisis de reconocidos investigadores epistemólogos y filósofos de dichos campos (Overton. [Developmental psychology](#), Valsiner. [A guided science](#)), el pensamiento psicológico moderno ha estado marcado por algunas dualidades centrales, tales como individuo/sociedad, mente/cuerpo, naturaleza/cultura, representación/realidad, hecho/valor, entre otros. Uno de nosotros ha propuesto que, en torno a la manera de tratar estas dualidades se pueden identificar dos ME en disputa (Castorina. [Las concepciones del mundo y los valores en la investigación psicológica](#); [Los supuestos filosóficos en psicología del desarrollo](#)).

Por un lado, se observa un ME hegemónico, heredero del pensamiento cartesiano, que tiende a disociar mente y cuerpo, representación y realidad, individuo y sociedad, denominado Marco Epistémico Escisionista (MEE, en adelante). Este MEE ha condicionado el desarrollo de los campos de investigación, constituyendo como unidad de análisis a un individuo disociado de sus condiciones sociales y de su contexto cultural, para centrarse exclusivamente en su procesamiento interno de la información. Esta delimitación ontológica es consistente con una metodología experimental que no va más allá de determinar y relacionar variables dependientes e independientes, sin mayor interés por los sentidos subjetivos o culturales, e infravalorando los métodos cualitativos. También se establecen los límites de la indagación cuando se promueve una psicología cerrada en sí misma, sin interés filosófico, u orientada hacia vínculos irreflexivos con las ciencias naturales, o sin diálogo con las ciencias sociales. Estos supuestos y diseños están orientados implícitamente por cursos de acción tales como la normalización de las conductas, controlar y optimizar comportamientos, o la promoción del individualismo. Así, en este programa, el ideal de la neutralidad valorativa del conocimiento científico aparece como un obstáculo epistemológico, una resistencia a cuestionar los usos de la teoría y los diseños metodológicos, asumiendo de antemano un recorte a priori de su



objeto de estudio. “Lo oculto opera sobre la formulación de los problemas hasta la justificación de sus hipótesis” (Castorina 2022:20).

Por otro lado, se propone un Marco Epistémico Relacional (MER, en adelante), que parte de un supuesto ontológico diferente, que supone que cada elemento de nuestra experiencia en el mundo existe en una conexión constitutiva con su opuesto: “la dimensión intersubjetiva conecta con la subjetiva, el organismo con el medio ambiente, la naturaleza con la cultura y el individuo con la sociedad” (Castorina 2021:6). Para el caso de las investigaciones centradas en el desarrollo del conocimiento, se asume que este es un proceso situado en contextos de prácticas y significados sociales preexistentes, y en relaciones con otros agentes sociales, de modo que se debe favorecer a una mirada amplia y dialógica. Los diseños metodológicos del MER favorecen técnicas que asumen que las unidades de análisis no son individuos-objetos, sino actores sociales reactivos a la situación de investigación, es decir, que reconocen y dialogan con la subjetividad del investigador. Estas definiciones son coherentes con valores sociales como la solidaridad, la promoción de la heterogeneidad y la diversidad cultural, la aspiración a transformar condiciones sociales injustas y el interés por dar voz a grupos sociales subordinados o marginados.

3. Dos marcos epistémicos en el campo del big data

En este apartado (y el siguiente) utilizamos el concepto de ME para ensayar una clave de análisis para los problemas epistemológicos del big data, poniendo así el foco en diversos aspectos en los que se expresa la relación conocimiento-sociedad. Para cumplir este objetivo, vamos a proponer dos tesis que suponen una cierta semejanza o analogía con lo observado en el campo de la psicología.

En primer lugar, partimos de constatar que la literatura epistemológica reciente del big data tiende a ordenar la discusión proponiendo una serie de dualidades o tensiones (Resnyansky. [Conceptual frameworks for social and cultural big data analytics](#); Leonelli. [Scientific research and big data](#); Wagner-Pacifi et al. [Ontologies, methodologies, and new uses of big data](#); Kitchin. [Big data, new epistemologies and paradigm shifts](#), entre otros). Por caso, uno de los principales temas de discusión es la naturaleza del dato y la pretendida objetividad de los análisis del big data, algo que ha sido problematizando como una disputa entre una visión representacional/relacional, o contraponiendo una concepción de dato crudo/interpretado, a la que podemos sumar la clásica dicotomía de hecho/valor, de larga presencia en la filosofía de las ciencias. Una segunda dimensión de la discusión es la que problematiza la manera de recortar su objeto de referencia, algo que observamos en una disputa entre conducta/sujeto. Vinculado también con la idea de objetividad, se encuentran otras dualidades en torno a cómo se caracteriza el análisis de los datos, que contraponen lo automático/no-automático, el aprendizaje supervisado/no-supervisado, o que distinguen entre inteligencia artificial/humana; muy cercano a esto se encuentran las discusiones sobre el fundamento o el modo de construir conocimiento, contraponiendo causalidad/correlación; y sobre lo que esperamos del big data, ya sea información/conocimiento. Tales dualidades no son las únicas que se discuten en la literatura. Por ejemplo, Resnyansky ([Conceptual frameworks for social and cultural big data analytics](#)) sugiere además que la retórica del big data ha operado entre la ubicuidad/visibilidad y que su alcance discurre entre aplicación universal/disciplinaria.

En segundo lugar, vamos a proponer que en torno a la manera en que se piensan o enforcan estas dualidades, podemos observar un ME tendiente a la escisión (MEE) y otro hacia la relación (MER).



El primero opta por reducir las dualidades a alguno de los términos, desconociendo o negando valor al otro, o defendiendo alguna forma de dualismo. El segundo parece asumir que los términos de las dualidades son indisociables y que deben ser pensados en una interacción constitutiva para lograr una mirada más compleja del fenómeno.

3.1 La naturaleza del dato

Uno de los primeros problemas corresponde a cómo entendemos la naturaleza del dato. Aquí Sabina Leonelli nos ofrece dos caracterizaciones que denomina la mirada representacional y la relacional: “La visión representacional interpreta los datos como representaciones confiables de la realidad que se producen a través de la interacción entre los humanos y el mundo. Las interacciones que generan datos pueden tener lugar en cualquier entorno social, independientemente de los fines de la investigación [versus] Desde el punto de vista relacional, los datos son objetos que se tratan como evidencia potencial o real para afirmaciones científicas de manera que pueden, al menos en principio, ser examinados y contabilizados. El significado asignado a los datos depende de su procedencia, sus características físicas y lo que estas características representan, y las motivaciones e instrumentos utilizados para visualizarlos y defender interpretaciones específicas” (Leonelli 2020:1). Análogamente, boyd y Crawford ([Critical questions for big data](#)) y Gitelman ([Raw data is an oxymoron](#)), proponen llevar la discusión de la pretendida objetividad del big data al contraponer la necesidad de la interpretación a la idea de un “dato crudo”.

La discusión en torno a la naturaleza de los datos no es abstracta. Una de las características centrales del big data es la movilidad o la vinculabilidad de los datos (Kitchin y McArdle. [What makes big data, big data?](#)), por un mayor grado de circulación, reutilización y entrecruce de los *datasets*, que resulta en una mayor distancia entre los contextos de registro y de interpretación. Frente a este escenario, una mirada representacional que escinde estos dos momentos puede seguir declamando la objetividad en base a la cantidad de dato disponible; mientras que una mirada preocupada por la relación entre los contextos no puede sino preguntarse hasta qué punto esta distancia no pone en riesgo la validez de la interpretación de los datos. Esto es más problemático cuando el contexto de registro y creación de los datos se sitúa en una plataforma digital o en sistemas de software, que implican una larga historia de decisiones técnicas y de pruebas de interacción con los usuarios, sujetas a cambios constantes, que no suelen ser registradas. Salganik ([Bit by bit](#)) ha documentado varios casos que señalan que el entramado de relaciones entre investigaciones y corporaciones que diseñan software es un desafío para la científicidad del conocimiento generado con big data, en tanto cualquier pretensión de validez estadística requiere de instrumentos de medida que produzcan datos de manera coherente y estable, a la vez que cualquier diseño experimental requiere control de variables y casos por parte del investigador, algo que las empresas son reticentes a ceder. En una mirada guiada por un MER, una deconstrucción de la historia del *dataset* es esencial para reconocer los límites a considerar en la interpretación.

Otra manera de problematizar el sentido epistemológico del dato es a través de la distinción hecho/valor. Esta es una dualidad de larga discusión en la filosofía de las ciencias, desde el momento en que el empirismo de Hume negó la posibilidad de inferir y justificar juicios morales a partir de afirmaciones de hecho; y desde que el positivismo lógico del siglo XX distinguió entre juicios falsables y analíticos (verdaderos en términos lógicos), dejando todo lo que no cumpliera esas condiciones en un contexto de afirmaciones “sin sentido”, fuera del análisis racional (Putnam. [The collapse of the fact/value dichotomy](#)). Esta dualidad ha sido profusamente documentada en el caso



del enfrentamiento entre el MEE y MER en psicología (Castorina. [Las concepciones del mundo y los valores en la investigación psicológica](#)). En el caso del big data, se discute la presencia de valores sociales tanto en la interpretación del dato como en su construcción. Esto último es fácil de rebatir en cuanto consideramos que gran parte de la datificación (el proceso de recolección, extracción, o apropiación de los datos) por parte de corporaciones se realiza en función de su valor económico. Así, “tanto en su producción como en su interpretación, todos los datos, incluidos los ‘big data’, son siempre el resultado de prácticas sociales contingentes y cuestionadas que brindan y ofuscan interpretaciones específicas del mundo” (Dalton y Thatcher 2014:5).

3.2 El objeto y el objetivo de estudio

Tal vez la manera más interesante para pensar cómo los valores sociales condicionan la construcción de conocimiento en el big data sea a través de la discusión acerca del recorte de su objeto de indagación (particularmente entre conducta/sujeto) y cómo esto se vincula con los objetivos sociales de la investigación en big data (generar explicaciones comprensivas o controlar conductas). Esta también es una discusión que ha tenido un largo desarrollo análogo en psicología y que uno de nosotros ha analizado proponiendo la oposición entre MEE y MER (Castorina. [The importance of worldviews for developmental psychology](#)): la primera posición corresponde a corrientes que entienden al individuo como aislado de un contexto social y que no consideran el sentido que adquiere la acción para el sujeto mismo, y a la vez, buscan un conocimiento que pudiera ser útil y eficaz en el control de su acción, en línea con el pragmatismo y el individualismo; la segunda adopta como unidades de análisis actividades y prácticas, y otras formas de relaciones entre individuo y su contexto sociocultural, y en ocasiones, adopta el objetivo de hacer explícitas situaciones de subalteridad, o dar voz a grupos oprimidos.

La discusión acerca de cómo entendemos y delimitamos a las entidades y agentes de los que habla el big data (social) se encuentra en constante redefinición a medida que avanza la datificación a ámbitos no alcanzados y se ponen en juego nuevas escalas y dimensiones de análisis. Ahora bien, es claro que la definición de los objetos de estudio del big data, a través de condicionamiento en la construcción del dato, se encuentran guiados por intereses y objetivos sociales. Zuboff nos ofrece una caracterización en línea con las características de un MEE: “Gracias a las capacidades de Big Other, el poder del instrumentista [del big data] reduce la experiencia humana a un comportamiento observable medible, mientras permanece firmemente indiferente al significado de esa experiencia” (Zuboff 2019:20). Una mirada alternativa podría ser la que, adoptando un MER, se oriente a observar el contexto a los sujetos y las prácticas, y a recuperar el sentido de sus acciones, para generar un sujeto más rico para la datificación, en lugar de su manipulación. Esta es la mirada que propone el proyecto del *Critical Data Studies* (sobre el que se cuenta con un monográfico en la revista [Big Data & Society](#)) y entre cuyos objetivos se incluye “exponer cómo los datos son intrínsecamente políticos y a los intereses a los que sirven; ... explorar cómo los nuevos regímenes de datos pueden ser utilizados de manera progresista” (Kitchin y Lauriault 2014:6).

3.3 El análisis de los datos

En relación con el momento del análisis podemos observar dos tendencias que disputan acerca de lo que implican los análisis automáticos, no-supervisados, que incluso, en ocasiones, se presentan como no-inteligibles. Generalmente, estas discusiones se han dado en torno al problema de la opacidad algorítmica, la dificultad para el analista de reproducir las capas de un cálculo



automatizado, y con ello, predecir sus resultados. Esto se puede deber a diversos factores, tales como la falta de acceso a los datos, parámetros y recursos computacionales; que la operatoria suponga una tarea de cálculo que excede a las posibilidades humanas; o que el algoritmo dependa de una historia, memoria o estado interno que lo vuelva una máquina “no trivial analíticamente indeterminable” (von Foerster 2003:311).

Ahora bien, el problema de la no-inteligibilidad y la opacidad algorítmica no puede ser un problema insalvable. Consideremos que otros criterios epistémicos y valores sociales (como la responsabilidad en la aplicación de un algoritmo, la deseabilidad del análisis automático en ciertos contextos o la renuncia a la búsqueda de explicaciones) dependen de que tomemos en serio este objetivo. Para ello es necesario asumir una mirada en línea con un MER que no limite la indagación de los algoritmos a su funcionamiento interno, sino más bien que observe cómo se insertan en entramados sociales. Esta es la propuesta de Kitchin ([Thinking critically about and researching algorithms](#)) dentro del mencionado *Critical Data Studies* o de la propuesta de Seaver ([Knowing algorithms](#)) por una etnografía de los algoritmos. Estos autores, sin obviar los intentos de “abrir” los algoritmos, coinciden en poner el foco en los grupos sociales involucrados en su producción (*deployment*), mantenimiento y uso.

Otra manera en que la literatura ha problematizado el análisis de los datos, es discutiendo el reclamo de algunos promotores del big data acerca de que la construcción de teorías y explicaciones causales se habría vuelto un objetivo obsoleto en la era del big data. En su lugar, estos promotores sostienen que generar conocimiento inductivamente a partir de correlaciones sería no solo suficiente, sino incluso más eficiente, ya que con la abundancia de datos el proceso es más barato y rápido y sería más fácil desde allí llegar a desarrollar predicciones estadísticas. Esta posición se puede discutir en varios niveles, comenzando por su dependencia en el impreciso (y también muy discutible empíricamente) supuesto de la mayor disponibilidad de datos. Brevemente: “más datos” no podrían asegurar una inferencia en tanto no aseguren la representatividad de los casos y que big data implique “todos los casos”, es directamente insostenible; de la misma manera “más datos” tampoco asegura que se cuente con los controles y las variables suficientes para desechar correlaciones espurias o para fundar una descripción externamente válida.

El planteo acerca de si el conocimiento requiere o no de postular explicaciones causales y construir teoría no es nuevo. La revalorización del modelado y otras tareas centradas en los datos se sustentan en una visión del conocimiento que lo vincula al mundo de la experiencia y las habilidades, en línea con el empirismo de Bacon o el pragmatismo de Dewey. Pero estas posiciones suponen un escepticismo que implica varias renunciaciones muy costosas para la misión social de la ciencia: la posibilidad de dar cuenta de un “por qué” que pueda generalizarse a nuevos escenarios; la posibilidad de hipotetizar sobre entidades que se pretenden parte de la realidad, a pesar de ser contraintuitivas; la posibilidad de actuar sobre esos nuevos espacios de realidad para transformarla; y finalmente, la posibilidad de proponer un proyecto colectivo racional, asentando en la idea de verdad como ideal regulatorio, sostenido por la reflexividad propia de la filosofía. Tal vez estas renunciaciones parezcan hoy aceptables por ser la expresión de la desvaloración (social) del conocimiento (científico) y un cierto reemplazo por la idea más limitada de “información”. De hecho, esto se sugiere a partir de los trabajos de uno de nosotros sobre la representación social del big data (Becerra, [Representaciones sociales del big data en la prensa digital argentina](#), Becerra y López-Alurralde, [Representaciones sociales del big data y la inteligencia artificial](#)), en las que las menciones a “dato” e “información” son mucho más salientes y centrales que la de “conocimiento”. La idea de



una jerarquía entre estos términos se puede rastrear en el modelo DIKW (influyente en *management*) que propone la pirámide datos > información > conocimiento > comprensión > sabiduría (Frické. [Data-information-knowledge-wisdom pyramid](#)). Aquí también la información es síntesis de los datos y el conocimiento es información que puede ser explotado en un contexto de uso, es decir, se trata de un conocimiento práctico, un *know-how*, antes que a un *know-that* proposicional.

Ahora bien, como señalan varios de los epistemólogos mencionados al inicio de este apartado, tal vez el problema sea la tendencia (que nosotros ubicamos en el fondo de un MEE) a pensar estos proyectos como alternativas u oposiciones y no como momentos o contextos que pueden integrarse (más en línea con un MER). En cualquier caso, la búsqueda inductiva de patrones y correlaciones puede ser muy útil para generar nuevas preguntas en un momento exploratorio, las cuales deben ser evaluadas a la luz de las teorías y los conocimientos previos de cada campo disciplinar. En términos de Clark y Golder: “Lo importante a tener en cuenta es que las ganancias de eficiencia en la recopilación de datos podrían, en principio, liberar recursos para la teoría formal y el diseño de investigación. Una cosa positiva, por lo tanto, sobre el cambio tecnológico es que significa un aumento en el poder adquisitivo, lo que significa que la competencia por los escasos recursos no tiene por qué ser de suma cero” (Clark y Golder 2014:66).

4. Una caracterización de los desafíos del big data desde un MER

En este último apartado queremos proponer una posible caracterización de los desafíos y oportunidades del big data para las ciencias (con foco en las ciencias sociales) desde una visión fundada en un marco epistémico relacional (MER). Esta sería una estrategia que busca proponer vínculos entre las dualidades planteadas ya que, en última instancia, se apoya en una cosmovisión que entiende que no puede escindirse el conocimiento científico y los valores sociales. Los desafíos a los que vamos a referirnos son:

1. Situar las oportunidades analíticas y metodológicas del big data en un diseño de investigación que proponga una dialéctica entre teoría y dato;
2. Incorporar una genuina reflexión y vigilancia epistemológica, considerando el contexto de relaciones sociales y económicas en el que se inscriben los momentos del registro y la recolección de los datos;
3. Promover la discusión racional en torno a los valores ético-políticos de la justificación y aplicación de la investigación con big data;
4. Y a partir de las discusiones epistemológicas que el big data está generando, promover una revisión de los problemas propios de las disciplinas que encaran el desarrollo de sus ramas computacionales.

En lo que queda de este último apartado queremos detallar estos desafíos, haciendo explícito los términos que vinculan o integran.

4.1 El uso del big data enmarcado y guiado por la teoría

La primera integración que proponemos apunta hacia una dialéctica entre construcción teórica y modelado a partir de los datos. Al igual que propone Kitchin, abogamos por una ciencia que, si lo requiere, “emplee técnicas cuantitativas, estadísticas inferenciales, modelado y simulación siendo



conscientes y abiertos a sus deficiencias epistemológicas, y recurriendo a la teoría social crítica para enmarcar cómo se lleva a cabo la investigación, cómo se da sentido a los hallazgos y el conocimiento empleado” (Kitchin 2014:9). En este sentido, se trata de reconocer que la exploración y el modelado de datos puede ser útil y valioso en distintos momentos del ciclo de una investigación, pero que esto dependerá de la manera en que se vincule con la teoría y el conocimiento de dominio del investigador. El desafío es, entonces, que causalidad y correlación (y deducción e inducción) se puedan integrar en diseños más creativos en los que los modelos explicativos y los procesos inferenciales implicados permitan generar exploraciones informadas y construcción de hipótesis y categorías a partir de la evidencia. Este proceso se deberá garantizar por un esfuerzo consciente de validación a través de múltiples técnicas y diversas formas de datos (incluyendo el *small data*) y de comunicación de las decisiones implicadas a lo largo del proceso. Así, tomados en su conjunto, creatividad, validación, transparencia e inteligibilidad podrían ser los vectores para un aprovechamiento del big data. Y, más importante aún, para retomar la pretensión de que se genere conocimiento vinculante (y no solo información) que nos pueda decir algo del mundo social y guiar nuestra praxis.

4.2 Reflexividad y vigilancia epistemológica en el big data

La segunda vinculación importante que se adopta como desafío es la que se da entre lo epistémico y lo social en el nivel de los datos. Para ello es importante reconocer la presencia y el condicionamiento de valores sociales a lo largo del proceso de construcción de conocimiento y la necesidad de una reflexividad y una vigilancia epistémica, capaces de tematizar y problematizarlos. Así, se impone como tarea una adecuada deconstrucción del dato, algo sobre lo que los científicos sociales estamos entrenados. En el caso del big data, se trata de indagar en torno a las prácticas de recolección vinculadas al negocio de *data warehousing* que pretenden una datificación “sin criterio ni sentido ... en la esperanza de que algún día estos datos asciendan a información” (Frické 2018:3). También supone analizar críticamente los modelos que se utilizan para el procesamiento, cuya construcción no suele ser accesible a los investigadores cuando se presentan como productos cerrados, escondiendo potenciales sesgos propios de procesos de reclutamiento o de la captura automática de datos (Lanier y Wiel. [AI is an ideology, not a technology](#)). Todo esto redundando en el desafío de que la vigilancia epistemológica no se limite al terreno del ciclo interno de la investigación, sino que considere también las relaciones entre el mundo académico y el comercial cuyas fronteras parecen estar borrándose por el impulso del big data. En última instancia, se trata de enfrentarnos a una pregunta difícil: ¿hasta qué punto la construcción de conocimiento científico empírico es compatible con una construcción del dato que se rige por la lógica (económica) de la acumulación y extracción?

4.3 Promover el debate sobre los fines del big data

Otra integración importante de lo epistémico y lo social es la que se da en la reflexión del tipo de objetivos de la investigación con big data y su vinculación con los intereses de los actores y las corporaciones, algo que nos resulta central para una crítica epistemológica guiada por la noción de ME. Aquí el riesgo es que se imponga institucionalmente un direccionamiento hacia formas limitadas de investigación, a través de regulaciones institucionales o mecanismos de incentivos. En términos de Cano: “las expectativas generadas por el empleo de datos masivos, podrían reforzar la presencia de una cierta forzosidad a la hora de elaborar modelos que ya no basta con que sean descriptivos sino que han de ser predictivos. Tal y como dice Taleb (2008), ‘la predicción está



firmemente institucionalizada en nuestro mundo' (p. 153), pero lo cierto es que la naturaleza de determinados fenómenos, como puedan ser el comportamiento humano ... son de naturaleza caótica o no lineal" (Cano 2016:465).

El desafío, en estos casos, es el de evitar caer en una "tiranía ontológica" que determine que el método (algorítmico) es el adecuado para el estudio de lo social y de lo humano, porque asume de antemano un recorte en su objeto de estudio (como puede ser la reducción de la experiencia a las conductas digitalmente medibles) a medida de ciertos intereses comerciales.

Otro camino que merece ser explorado es el de la integración de los actores sociales en la construcción de los modelos, en la discusión de los intereses de la investigación y hasta de las condiciones éticas en las que se desarrolla. En este tema tal vez nos sirvan las experiencias aprendidas en el campo de la sociocibernética y el estudio de los sistemas sociales complejos de los años 80, un antecedente de las ciencias sociales en su apertura al diálogo interdisciplinario y al uso de modelos computacionales (Becerra. [Sociocibernética](#)). Allí, por caso, se aceptan diversas formas de modelado como instrumento metodológico; y en algunos proyectos se ha ensayado modelado participativo con los actores sociales involucrados, como forma de incluir sus prioridades en la construcción y calibración de los desarrollos y en la discusión de objetivos y escenarios sociales alternativos. Más recientemente, Salganik ([Bit by bit](#)) propone entre las distintas oportunidades del big data la de diseñar experimentos digitales e incentivar la colaboración masiva del público. Un diseño muy interesante (y cada vez más accesible) es el de crear entornos digitales (*software* o *apps*) con un fin que motive a los participantes, que aproveche su heterogeneidad y que permita hacer un uso ético de los datos.

4.4 Revisión hacia y desde dentro de los problemas disciplinares

Una última línea de revinculación que proponemos explorar es la que se da en torno a dicotomías propias de los campos disciplinares sociales que, con la incorporación del big data, avanzan hacia la integración de lo computacional y que se pueden servir de las oportunidades de articulación entre automático/no-automático, la inteligencia artificial/humana o lo material/digital, para repensar algunas dicotomías y dualismos propios que han marcado su desarrollo. Nos referimos, en el caso de la sociología, a dualismos como cuantitativo/cualitativo, micro/macro, laboratorio/campo, hermenéutica/explicación y (conectado con lo tratado en el párrafo anterior) entre académico/privado. Es decir, el desafío consiste en pensar al big data como una oportunidad, pero no por las razones que promulga su retórica, que son ajenas a los intereses y necesidades de las ciencias sociales, sino por su potencialidad para visitar algunos de los debates internos que han marcado la propia historia de estos campos.

Por caso, pensemos en el primero de estos dualismos: cuantitativo/cualitativo. Como varios autores señalan (Diaz-Bone et. al. [Social research in times of big data](#)) creer que el big data se limita a reificar el ánimo cuantitativo de las ciencias sociales sería un error ya que, como mínimo, pone en cuestión algunos estándares de estas metodologías, como la consideración muestral o el modelo de hipótesis nula. De hecho, como hemos propuesto en otro trabajo (Becerra y López-Alurralde. [Topic modeling y los desafíos de la investigación cualitativa](#)) en relación a una técnica difundida de análisis textual que se incluye en el campo del *machine learning* no-supervisado o automático, denominada *topic modeling*, los problemas metodológicos son análogos a los de la investigación cualitativa: las decisiones y parámetros adoptados en estos procesos algorítmicos se condicionan iterativamente,



de modo que la validez dependerá, en última instancia, de ofrecer una documentación transparente, una interpretación razonable y creíble y formas de validación replicable que permitan la discusión de nuestras inferencias y el planteo de alternativas. Y esto sin mencionar los conocidos debates éticos en torno a diversas tareas de investigación, como es el manejo de fuentes y el cuidado de la privacidad en la era del big data.

Conclusiones

En este trabajo nos propusimos indagar el discurso epistemológico en torno al big data, apelando a la noción de marco epistémico (ME) del constructivismo piagetiano revisado, cuya utilidad reside en poner el foco en los distintos supuestos epistémicos y valores no epistémicos que, indirectamente y sin determinarlos, condicionan la construcción de conocimiento. Particularmente nos centramos en una utilización del ME que lo entiende como una cosmovisión que se refleja en un posicionamiento particular del vínculo conocimiento-sociedad y que alcanza a las formas de interpelar a los objetos y de involucrarse en el proceso de construcción del conocimiento, que incide en las preguntas y los objetivos últimos, el alcance de la reflexión, la selección de métodos y la práctica de los investigadores. Un elemento central de estas cosmovisiones es su adherencia a valores e intereses sociales que condicionan la investigación desde su concepción, justificación y aplicación y que pueden resultar en una problematización limitada de la realidad social a través del big data o habilitar la tematización sobre las relaciones sociales implicadas en la producción de conocimiento con big data.

En los apartados anteriores hemos extendido las exploraciones que se han realizado en otros campos del conocimiento (como la psicología del desarrollo y la psicología social) hacia la reflexión de los ME comprometidos en el big data. Estos intervendrían en la manera en que se enfocan algunas dualidades, tales como causalidad/correlación, dato representacional/relacional, inteligencia artificial/humana o hecho/valor. Luego de revisar los debates recientes en torno a algunas de estas dualidades, hemos realizado una caracterización de las oportunidades y desafíos del big data (con mayor foco en las ciencias sociales) desde un MER. Nuestra propuesta se enfoca en cuatro desafíos: la búsqueda de diseños que propongan una relación virtuosa entre teoría y dato, evitando sesgos empiristas que resulten en renuncias en el proyecto de un conocimiento vinculante y socialmente comprometido; la necesidad de profundizar la reflexión y la vigilancia epistemológica en la relación conocimiento-sociedad, que nos compele a ver al big data como un fenómeno que cruza los límites de lo académico y que introduce condicionamientos normativos e institucionales en la construcción del conocimiento que deben ser puesto en tela de juicio crítico; la necesidad de promover la discusión en torno a los valores sociales en la construcción, justificación y aplicación de la investigación con big data, así como discutir el rol de actores y corporaciones; y en la adopción de criterios para evaluar las oportunidades del estudio del big data y el avance de lo computacional en las ciencias sociales, que se orienten a enfrentar algunos de los desafíos registrados en la historia de este campo, de forma tal que se propicie una reapropiación del big data. Creemos que esta orientación que nos propone el MER cumple con el desafío que reclama Kitchin (2014) acerca de una epistemología *situada, reflexiva y contextualmente vinculada* para un aprovechamiento fructífero del big data.

Con vistas al desarrollo de esta visión epistemológica, se abre un diálogo con otros programas epistemológicos. Uno de ellos puede ser con la noción de “paradigma” de Kuhn, que fue la principal inspiración de Piaget y García para elaborar la noción de ME. Nosotros creemos que un paso natural



sería el de acercar este enfoque a la discusión sobre si el big data constituye un nuevo paradigma científico. También se impone incorporar críticamente aspectos de otros programas epistemológicos que se han ocupado de la relación entre conocimiento (científico) y sociedad, como pueden ser los trabajos de Douglas ([Rejecting the ideal of value-free science](#)) o Longino ([Science as social knowledge](#)), que han logrado revitalizar la discusión acerca de la objetividad en base al tratamiento del lugar de los valores sociales en la producción científica.

Bibliografía

- boyd, D. y Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, 15(5), 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
- Cano, I. (2016). Reflexiones epistemológicas sobre big data. *Eikasia. Revista de Filosofía*, 71, 449-474. <https://dialnet.unirioja.es/servlet/articulo?codigo=6855938>
- Castorina, J.A. (2022). Los supuestos filosóficos en psicología del desarrollo: ¿Meta teoría o marco epistémico? *Revista de Psicología*, 21(2), 2-28. <https://doi.org/10.24215/2422572Xe140>
- Castorina, J.A. (2021). The importance of worldviews for developmental psychology. *Human Arenas*, 4, 153-171. <https://doi.org/10.1007/s42087-020-00115-9>
- Clark, W.R. y Golder, M. (2014). Big data, causal inference, and formal theory: contradictory trends in political science? *Political Science and Politics*, 48(1), 65-70. <https://doi.org/10.1017/S1049096514001759>
- Dalton, J. y Thatcher, J. (2014). What does a critical data studies look like, and why do we care? *Society and Space* <http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher>
- Frické, M.H. (2018). Data-information-knowledge-wisdom (dikw) pyramid, framework, continuum. *Encyclopedia of Big Data*. https://doi.org/10.1007/978-3-319-32001-4_331-1
- García, R. (2000). *El conocimiento en construcción. De las formulaciones de Jean Piaget a la teoría de sistemas complejos*. Gedisa.
- Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1) <http://journals.sagepub.com/doi/10.1177/2053951714528481>
- Kitchin, R. y Lauriault, T.P. (2014). Towards critical data studies: charting and unpacking data assemblages and their work, pp. 1-19. In J. Eckert et al. *Geoweb and big data*. University of Nebraska Press.
- Leonelli, S. (2020). Scientific research and big data. *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/science-big-data>
- von Foerster, H. (2003). *Understanding understanding: essays on cybernetics and cognition*. Springer.
- Zuboff, S. (2019). Surveillance capitalism and the challenge of collective action. *New Labor Forum*, 28(1), 10-29. <https://doi.org/10.1177/1095796018819461>

Recibido el 21 nov 2022

Aceptado el 4 ene 2023