

Ciencia de datos y estudios globales: aportaciones y desafíos metodológicos

Daniel Lemus-Delgado
Tecnológico de Monterrey (México)

Ricardo Pérez Navarro
Tecnológico de Monterrey (México)

CÓMO CITAR:

Lemus-Delgado, Daniel y Ricardo Pérez Navarro. 2020. "Ciencias de datos y estudios globales: aportaciones y desafíos metodológicos". *Colombia Internacional* (102): 41-62. <https://doi.org/10.7440/colombiaint102.2020.03>

RECIBIDO: 22 de agosto de 2019

ACEPTADO: 13 de noviembre de 2019

MODIFICADO: 29 de enero de 2020

<https://doi.org/10.7440/colombiaint102.2020.03>

RESUMEN. **Objetivo/Contexto:** La ciencia de datos es considerada un enfoque novedoso y promisorio empleado en la obtención y análisis de información en múltiples disciplinas científicas. A pesar de su amplia difusión y aceptación, el uso de técnicas basadas en la ciencia de datos en el campo de los estudios globales está en ciernes. En este contexto, este artículo tiene como objetivo discutir las contribuciones y desafíos metodológicos que la ciencia de datos puede aportar a la disciplina de los estudios globales. **Metodología:** Este artículo parte del análisis crítico de información sobre los fundamentos conceptuales de la ciencia de datos y lo contrasta con sus posibles aplicaciones en la disciplina de los estudios globales. De esta manera, se examinan los elementos conceptuales que definen la ciencia de datos, se establecen las concepciones paradigmáticas de los estudios globales, se analizan los vínculos posibles entre la ciencia de datos y los estudios globales y se discuten los límites y alcances que la ciencia de datos puede aportar como enfoque metodológico. **Conclusiones:** Este artículo concluye que las herramientas tecnológicas basadas en *big data* pueden enriquecer nuestra comprensión sobre los fenómenos globales, siempre y cuando se asuma una actitud crítica que reconozca que tanto la elección de los datos como su análisis están embebidos en contextos históricos y sociales. Por lo tanto, como cualquier otro enfoque, la ciencia de datos representa una manera parcial de capturar y analizar información. **Originalidad:** La originalidad de este artículo recae en asumir una reflexión crítica sobre los alcances y límites de la ciencia de datos, más allá de una visión positivista del conocimiento científico, reconociendo los sesgos que existen detrás de las explicaciones supuestamente neutrales que encierra el empleo de técnicas de investigación basadas en la ciencia de datos.

PALABRAS CLAVE: Estudios globales; ciencias de datos; investigación científica; metodología.

Data Science and Global Studies: Contributions and Methodological Challenges

ABSTRACT. Objective/Context: Big Data is a novel and promising methodological approach to acquire and analyze information from an extensive range of scientific disciplines. Despite the fact that Big Data has wide dissemination and acceptance in different scientific disciplines, its use in the field of Global Studies is still incipient. Taking this into account, this paper discusses the methodological challenges and presumable contributions that Big Data offers to Global Studies. **Methodology:** The starting point for this paper is the critical analysis of the conceptual bases of Big Data and its relationships with Global Studies. So, this paper proposes the possible applications of Big Data in the Global Studies contrasting some key concept. In this way, the conceptual elements that define Data Science are examined and compared with the paradigmatic conceptions of Global Studies. On these conceptual fundamentals, the article discusses links, limits and, scopes of Big Data and how it can contribute to Global Studies. **Conclusions:** This paper concludes that technological tools based on Big Data can enrich our understanding of global phenomena only if we position ourselves in a critical attitude that recognizes that the choice of data and its analysis are decisions embedded in historical and social contexts. In this way, as with any other methodological approach, Big Data is only a partial way to obtain and analyze information. **Originality:** The originality of this paper is to shed critical light on the scope and limits of Big Data, beyond a positivist vision of scientific knowledge. In other words, this article recognizes the biases behind the supposed neutral explanations regarding the use of techniques based on Big Data.

KEYWORDS: Global studies; big data; scientific research; methodology.

Ciência de dados e estudos globais: contribuições e desafios metodológicos

RESUMO. OBJETIVO/CONTEXTO: A ciência de dados é considerada uma abordagem nova e promissora, utilizada por várias disciplinas científicas para a obtenção e análise de informações. Apesar de sua ampla disseminação e aceitação, o uso de técnicas baseadas na ciência de dados no campo de estudos globais está em estágio inicial. Nesse contexto, este artigo tem como objetivo discutir as contribuições e desafios metodológicos que a ciência de dados pode trazer para a disciplina de estudos globais. **Metodologia:** Este artigo analisa criticamente os fundamentos conceituais da ciência de dados e os compara com as possíveis aplicações na área de estudos globais. Desse modo, este artigo examina os elementos conceituais que definem a ciência de dados, estabelece os paradigmas conceituais dos estudos globais, analisa as possíveis ligações entre ciência de dados e estudos globais, e discute como os limites e alcance da ciência de dados pode contribuir como abordagem metodológica. **Conclusões:** este artigo conclui que as ferramentas tecnológicas baseadas em *big data* podem enriquecer nossa compreensão sobre os fenômenos globais, desde que seja assumida uma atitude crítica que reconheça que a seleção dos dados e sua análise estão incorporadas em contextos históricos e sociais. Deste modo, como em qualquer outra abordagem, a ciência de dados representa um modo tendencioso de capturar e analisar

informações. **Originalidade:** a originalidade deste artigo consiste em assumir uma reflexão crítica sobre o alcance e os limites da ciência de dados, além de fornecer uma visão positivista do conhecimento científico, reconhecendo a parcialidade que existe por trás das explicações supostamente neutras sobre o emprego de técnicas de pesquisa baseadas em ciência de dados.

PALAVRAS CHAVE: Estudos globais; ciências de dados; Pesquisa científica; metodologia

Introducción: las tendencias globalizantes y la ciencia de datos

El concepto de lo *global* se ha convertido en un término utilizado ampliamente en el análisis de múltiples matices de la realidad contemporánea. Las transformaciones de gran alcance que se experimentan mundialmente plantean la necesidad de nuevos paradigmas que permitan comprender mejor las dinámicas de dichas transformaciones. Es en este contexto en el que se origina la disciplina de los estudios globales con el propósito de comprender mejor el impacto que las tendencias de la globalización tienen sobre las distintas facetas de nuestro entorno.

Sin lugar a dudas, los cambios que experimentamos hoy en día están modificando el mapa que permite organizar nuestro conocimiento sobre la realidad. Al mismo tiempo, debido al acceso masivo de datos como nunca antes había ocurrido, se experimenta un profundo cambio sobre las posibilidades de conocer. En otras palabras, simultáneamente presenciamos dos profundas transformaciones. Por una parte, “el territorio” que intentamos reflejar en un mapa se ha modificado en las últimas décadas como resultado de que el mundo es cada día más estrecho y está más interconectado social, política y económicamente. Por otra parte, la cantidad de información a nuestro alcance para elaborar dicho mapa ha crecido exponencialmente. Esta es una transformación dual tanto de lo ontológico como de lo epistemológico. De este modo, dichos cambios están modificando al mismo tiempo la esencia de la realidad y la manera en que se obtiene la información. Bajo esta perspectiva, este artículo explora los alcances y límites de las nuevas herramientas metodológicas basadas en la ciencia de datos y su posible impacto en la disciplina de los estudios globales.

En efecto, una de las transformaciones más importantes que están ocurriendo en el presente siglo es el crecimiento exponencial de datos que abre la posibilidad a nuevas formas para el análisis de la realidad. En los últimos años hemos presenciado la aparición de *big data*, tecnologías disruptivas que han modificado considerables aspectos de la sociedad (Shan *et al.* 2015, 39). Como lo han observado Mayer-Schönberger y Cukier (2013, 3), hasta hace poco tiempo la

información era escasa, costosa y difícil de conseguir. No obstante, la expectativa era que esa información fuera de una calidad lo suficientemente elevada de modo que permitiera obtener conclusiones que resultaran razonablemente válidas. Hoy en día, sin embargo, con la recopilación y almacenamiento de datos masivos de información, se está modificando la manera en que se aprehende la realidad. Este hecho supone un drástico cambio de enfoque en la forma de comprender el mundo. El valor de la información ya no reside en datos concretos, sino en la forma en que se correlacionan datos masivos para descubrir patrones que ni siquiera se habían imaginado. Así, cuando se analizan computacionalmente datos masivos, se puede proporcionar información más precisa sobre patrones, tendencias y asociaciones ocultas (Grable y Lyons 2018, 17). Por lo tanto, es posible sugerir que vivimos en la época de la revolución de los datos. Se trata de un mundo en el que estos son más grandes, rápidos y detallados que nunca, como consecuencia de las nuevas tecnologías que están propiciando un aumento exponencial con respecto al volumen y los tipos de datos disponibles, y crean así posibilidades sin precedentes para informar y transformar la sociedad (Naciones Unidas 2014, 2).

Hasta ahora, el análisis de datos masivos se ha utilizado en distintas disciplinas de investigación científica cuya finalidad ha sido resolver problemas complejos. Estas investigaciones incluyen, entre otros campos, estudios sobre medio ambiente, educación, salud, transporte, seguridad nacional y biomedicina (Simpson 2016, 8). Entre estos proyectos destacan el uso de técnicas de extracción de datos para mejorar y predecir la empleabilidad (Saouabi y Ezzati, 2019, 412). En la agricultura, la ciencia de datos se ha usado para elevar la productividad agrícola mediante la recopilación y el procesamiento de información sobre el crecimiento de las plantas, el monitoreo de gases de efecto invernadero y el análisis de las posibles consecuencias del cambio climático en los cultivos (Mathivanan y Jayagopal 2019, 55). Además, las técnicas de *big data* han facilitado muchos de los avances recientes en el campo de la medicina de precisión y se tiene la expectativa de que sustenten la mayoría de los adelantos en el tratamiento y la prevención de enfermedades que aún están por venir (Clare y Shaw 2016, 1). Asimismo, el *big data* ha sido utilizado en el estudio de la comunicación verbal y no verbal, tanto con métodos computacionales como cognitivos; en el desarrollo de The Cancer Genome Atlas, para apoyar las investigaciones relacionadas con el cáncer; en el análisis de múltiples secuencias de proteínas, para determinar los enlaces evolutivos y predecir estructuras moleculares; en el desarrollo de sistemas de inteligencia territorial de ciudades, para potencializar la calidad de vida de los habitantes de las grandes urbes, y en el monitoreo de bebés prematuros en las salas de neonatología, para determinar cambios en la presión arterial o la temperatura, y alteraciones en los registros de los electrocardiogramas y electroencefalogramas (Barranco 2012).

Sin embargo, el análisis de los datos masivos en el campo de los estudios globales es un tema aún poco explorado. Este artículo contribuye al debate sobre las aportaciones que los estudios globales pueden realizar a la comprensión de la complejidad y diversidad de un mundo globalizado, a partir de las oportunidades y limitaciones metodológicas que la ciencia de datos ofrece. Para ello, se organiza en cuatro secciones. En la primera se discuten teóricamente cuáles son los elementos que definen la ciencia de datos. En la segunda sección se relacionan las posibilidades que ofrece la ciencia de datos para los estudios globales. Más adelante, se discuten las implicaciones teórico-metodológicas del uso masivo de datos para los estudios globales. Por último, en las conclusiones se destacan los límites y posibilidades de la ciencia de datos en la comprensión de las tendencias globales.

1. Una aproximación conceptual a la ciencia de datos

No es posible hacer una aproximación conceptual a la ciencia de datos sin tener en cuenta el surgimiento del Internet, ya que esta tecnología representa un cambio radical en la manera en que la sociedad genera y utiliza la información. El Internet puede ser definido como una red de computadoras que se encuentran interconectadas, a nivel mundial, con la finalidad de compartir información. En este sentido, se trata de una red compleja que conecta una gran cantidad de dispositivos, como computadoras, servidores de archivos y cámaras de video, a través de líneas telefónicas, satélites y redes inalámbricas (Araya 2005, 1049). Esta tecnología fue el resultado de la guerra fría, cuando los Estados Unidos y la Unión Soviética se lanzaron a una carrera frenética por alcanzar la supremacía mundial, que incluía también el ámbito tecnológico. De esta manera, la Unión Soviética lanzó en 1957 el primer satélite artificial Sputnik. Como respuesta, los Estados Unidos establecieron la Advanced Research Projects Agency. Al final del día, esta agencia fue la que estableció los cimientos para el desarrollo del Internet. Así, en 1962, Paul Baran presentó un sistema de comunicaciones que, mediante computadoras conectadas a una red descentralizada, resultaba inmune a ataques externos (Mieczkowski 2013, 4).

Este sistema de comunicaciones fue evolucionando mediante la integración a la red virtual, en primer lugar, de los sitios de las universidades; más tarde, se incorporarían plataformas de negocios, se desarrollarían las redes sociales y se generaría una serie de aplicaciones, como la georreferenciación. Esta conexión virtual es la que ha generado un continuo y creciente cúmulo de información (Ryan 2009, 23-30). Se puede afirmar, metafóricamente hablando, que pasamos de ser individuos análogos a individuos digitales, lo que ha propiciado que ahora estemos permanentemente conectados desde diferentes dispositivos prácticamente

en todo momento. Esto genera una enorme cantidad de datos que pueden ser obtenidos, clasificados y analizados gracias al avance de computadoras que cada día son más robustas (Hilbert y López 2011, 60-65).

En un primer momento, la generación en grandes volúmenes de información fue aprovechada por empresas especializadas en la administración de sitios de Internet con el objetivo de encontrar en ese conjunto de datos oportunidades para incrementar sus utilidades (Van der Aalst 2014, 13-28). Sin embargo, este nuevo contexto caracterizado por la generación masiva de información ha dado paso al establecimiento de una sociedad basada en el conocimiento, donde se destaca una nueva profesión, la de los *trabajadores del conocimiento*, quienes cuentan con capacidades, habilidades y el acceso a la tecnología que les permite procesar el continuo flujo de información (Drucker 1993). Este trabajador del conocimiento ha evolucionado hacia un nuevo perfil de empleado, que actualmente se conoce como *científico de datos*. Este profesional combina la estadística, las matemáticas, la programación y las habilidades para la solución de problemas con la captura datos, y es capaz de llevar a cabo las actividades propias de limpieza, preparación e integración de datos con la intención de localizar patrones en la información analizada (Davenport y Patil 2012, 5-7).

Ahora bien, la ciencia de datos se compone de tres áreas. La primera es el *big data*, que se emplea para procesar los datos. La segunda es la minería de datos, cuya finalidad es encontrar patrones, incluso sin que estos fueran antes imaginados. Por último, la visualización de los datos, cuyo propósito es facilitar la comprensión de la información de manera clara y propiciar su socialización.

El concepto de *big data* fue empleado por primera vez por Cox y Ellsworth (1997, 5-7), quienes lo definieron como el proceso de graficar, para su correcta visualización, grandes cantidades de información, dadas las limitaciones técnicas de las computadoras. Lafrate (2018, 77-82) señala que *big data* son datos que contienen una mayor variedad que los comunes, que se presentan en volúmenes crecientes, a una gran velocidad y que poseen valor al ser inteligentes.

Desde una perspectiva crítica, Boyd y Crawford (2012, 663) definen *big data* como un fenómeno cultural, tecnológico y académico que se basa en la interacción de la tecnología —que maximiza la potencia de cálculo y la precisión algorítmica para recopilar, analizar, vincular y comparar grandes conjuntos de datos—, el análisis —que aprovecha grandes conjuntos de datos para identificar patrones— y la mitología —que asume la creencia de que los grandes conjuntos de datos ofrecen una forma superior de inteligencia y conocimiento que puede generar ideas que antes eran imposibles, con un aura de verdad, objetividad y precisión—.

Para tener una aproximación a lo que significan los datos masivos, se puede señalar que, desde marzo del año 2018, la red social de Facebook tiene

registrados 2.200 millones de usuarios activos mensuales y 1.450 millones de usuarios activos diariamente, cifra que representa un 13 % más que el año anterior; asimismo, se estima que existen 1.500 millones de usuarios activos mensuales de WhatsApp; además, diariamente se envían alrededor de 500 millones de *tweets* en la plataforma de Twitter, lo que significa que en los últimos seis años aumentó diez veces; por último, se estima que diariamente se realizan más de 7.000 millones de consultas en el buscador de Google (Gupta y Rani 2019, 322).

En este sentido, Nunan y Di Domenico (2013, 3-6) establecen tres retos que los científicos sociales deben tener en cuenta en relación con el fenómeno del *big data*. Primero, los problemas tecnológicos asociados al almacenamiento, la seguridad y el análisis de crecientes volúmenes de datos. Segundo, el valor comercial que puede ser añadido a través de la generación de percepciones más efectivas. Finalmente, los impactos sociales, particularmente las implicaciones que el uso de datos tiene para la privacidad personal.

Así, ante la enorme cantidad de datos que se generan cada día, Van der Aalst (2014, 13-28) propone un marco de referencia para identificar el proceso de generación de información de acuerdo con el término *Internet of events* (IoE), con la intención de clasificar todos los datos disponibles en Internet. Así, se establecen tres categorías. La primera es el *Internet of the content* (IoC), que representa la información generada por personas que buscan aumentar el conocimiento sobre temas particulares; ejemplos de esta información son artículos y blogs, enciclopedias como Wikipedia, y plataformas de video como YouTube y de libros electrónicos como Google Books. La segunda, *Internet of the people* (IoP), incluye la información generada por la interacción social; por ejemplo, el correo electrónico, las redes sociales y los foros virtuales. La tercera categoría la comprende el *Internet of the things* (IoT), la cual se compone por la información generada por los objetos conectados a la red; se trata de todas las cosas que tienen una identificación única y una presencia en una estructura similar a Internet. En este sentido, las cosas pueden tener una conexión a Internet o estar etiquetadas a la red. Por último, se encuentra el *Internet of the locations* (IoL), que comprende la información que cuenta con una dimensión espacial derivada de la adopción de dispositivos móviles; por ejemplo, los teléfonos inteligentes que generan cada vez más eventos con atributos geoespaciales.

En lo que respecta a la segunda área del *big data*, referente a la minería de datos o *data mining*, es preciso destacar que no hay una definición única. Hand, Mannila y Smyth (2001, 1) la consideran como el análisis de conjuntos de datos observados, generalmente de gran volumen, con el objetivo de encontrar nuevas relaciones entre variables, así como el correcto resumen de dichos conjuntos

de datos en forma comprensible y útil. Para Graham (2011, 2-5), *data mining* es el arte y la ciencia del análisis inteligente de datos, con el objetivo de generar conocimiento de interés a partir de estos. Por otra parte, la minería de datos se considera un paso fundamental en el descubrimiento de patrones de información o *knowledge discovery*. Este se entiende como el proceso de encontrar patrones válidos relevantes, potencialmente útiles, buscando que sean comprensibles (Klösigen y Zytchow 2002, 10-21).

Al ser la minería de datos un proceso para extraer información, esta recurre a métodos y modelos que permitan la definición de relaciones o patrones entre variables no conocidos previamente. Ahora bien, los modelos que emplean son de dos tipos: los descriptivos, que buscan identificar patrones que permitan resumir y explicar el comportamiento de los datos; y los modelos predictivos, los cuales buscan estimar los valores futuros de una variable de interés con base en su comportamiento histórico (Riquelme, Ruiz y Gilbert 2006, 12). En conjunto, el proceso de la minería de datos incluye la descripción, la estimación, la predicción, la clasificación, el agrupamiento y la asociación de datos (Larose y Larose 2014, 8).

Chitra y Subashini (2013, 220-222) clasifican las técnicas que se emplean en estos modelos como supervisadas y no supervisadas. Las técnicas supervisadas se utilizan para construir modelos que son empleados para realizar predicciones. Entre estas destacan el desarrollo de algoritmos genéricos, redes neuronales, arboles de decisión y análisis de regresión. Las técnicas no supervisadas, o algoritmos de descubrimiento del conocimiento, se usan generalmente para la extracción de información útil a partir de grandes volúmenes de datos. Ejemplos de este tipo son el *clustering*, el análisis de enlace y el análisis de frecuencia.

El último elemento de la ciencia de datos es la visualización. Esta técnica permite examinar una gran cantidad de datos e identificar patrones o tendencias con la ayuda de gráficas o representaciones, empleando diferentes métodos y técnicas (Ware 2012, 1-9). Si bien es cierto que la visualización de la información ha estado presente en el desarrollo de la humanidad, solo recientemente su uso se ha extendido gracias al *software* de mayor capacidad de procesamiento y al desarrollo de las llamadas librerías que permiten graficar y representar de forma fácil datos en mapas y tableros de control.

Evidentemente, cada uno de los componentes de la ciencia de datos se encuentra en continua evolución. Sin duda, este hecho influye en la forma en que se lleva a cabo la investigación científica e impacta en diferentes áreas del conocimiento. Así, en la siguiente sección se analizarán las posibilidades de una relación promisoría entre los estudios globales y la ciencia de datos.

2. Los estudios globales y la ciencia de datos: ¿una relación promisoria?

A pesar de que el fenómeno de lo global no es una tendencia exclusiva de la vida contemporánea, los estudios globales son una disciplina científica relativamente nueva, ya que solo en tiempos recientes los científicos sociales comenzaron a analizar sistemáticamente redes, flujos, procesos, ideologías y representaciones de los sistemas transnacionales y globales, tanto desde una perspectiva histórica como desde una aproximación contemporánea (Juergensmeyer 2013, 4). El campo académico de los estudios globales surgió en el actual contexto de una creciente globalización, como respuesta a la búsqueda por comprender mejor los cambios que se han intensificado en las últimas décadas y que han incrementado las interconexiones políticas, económicas y sociales a escala planetaria (Campbell, MacKinnon y Stevens 2010, 2). Desde que, en la década de 1990, se establecieron los primeros programas académicos y centros de investigación sobre los estudios globales en Norteamérica, Europa y Asia, esta disciplina ha florecido ampliamente. Actualmente, se estima que existen alrededor del mundo aproximadamente doscientos centros de investigación que abordan el fenómeno de lo global desde diferentes ángulos (Juergensmeyer 2012, 849-857).

Sin embargo, cuando se parte desde la perspectiva de los estudios globales, es necesario tener en cuenta que esta disciplina no significa solamente analizar la globalización. Al respecto, Nederveen (2013) propone tres aspectos en que los estudios globales se distinguen con claridad de los estudios sobre la globalización. Primero, sugiere que los estudios globales surgen de una ola cualitativamente diferente de aquella que impulsó el análisis de la globalización desde disciplinas específicas que cuentan ya con una larga tradición intelectual. Es decir, los estudios globales no abordan asuntos referentes a disciplinas específicas incorporando lo global como un tema de estudio, como sucede en el caso de la *economía política global*, la *cultura global* o la *comunicación global*. En este sentido, propone que primero se intensificaron las tendencias globales; después, la globalización empezó a ser estudiada desde campos del saber ya establecidos; por último, hoy en día nos encontramos en una tercera fase en que la disciplina de los estudios globales surge como una forma distinta de analizar la globalización. Basándose en un enfoque interdisciplinario, esta manera de abordar lo global rebasa las fronteras tradicionales del saber contenidas en disciplinas específicas. Segundo, advierte que, desde una perspectiva intelectual, los estudios globales son aún una disciplina en ciernes, a lo que él llama *andamiajes sin techo*. Tercero, afirma que los estudios globales pueden agregar valor más allá de los tradicionales estudios internacionales, porque aspiran a una construcción del conocimiento multicéntrica, que no

tiene solamente como centro de conocimiento y experiencia histórica al mundo occidental, y parte de una perspectiva multinivel, en la que lo local, lo nacional y lo mundial tienen la misma importancia (Nederveen 2013, 503-513).

Al definir los estudios globales, es posible destacar las siguientes características fundamentales. En primer lugar, son transnacionales, ya que se centran en el análisis de eventos, actividades, ideas, tendencias, procesos y fenómenos que aparecen a través de las fronteras nacionales y las regiones culturales. Además, son interdisciplinarios, debido a que los fenómenos que abordan son económicos, políticos, sociales, culturales, religiosos, ideológicos, ambientales y biológicos, y, por lo tanto, se examinan desde distintos horizontes disciplinarios. Asimismo, los estudios globales son a la vez históricos y contemporáneos porque, para comprender plenamente los actuales patrones de lo global, es necesario analizar sus precedentes históricos. Finalmente, tienden a ser poscoloniales y críticos. Aunque muchos aspectos de la globalización contemporánea se basan en precedentes coloniales europeos, generalmente los estudios globales no aceptan sin crítica los patrones de la globalización económica, política y cultural que se fraguaron desde Occidente. En este sentido, reconocen que los problemas, dinámicas y tendencias globales se aprecian de manera diferente desde distintas partes del mundo y, dentro de un mismo país o región, desde diversas ubicaciones, dependiendo del lugar social y económico que ocupen quienes perciban estos fenómenos. Por lo tanto, los estudios globales proponen hablar de *múltiples globalizaciones*, reconociendo que no hay un paradigma o perspectiva dominante que se valore más que otros (Juergensmeyer 2013, XIII-XVII).

De este modo, el giro colectivo de diversas disciplinas para involucrarse en la comprensión de los procesos de la globalización y sus problemas inherentes representa algo más que una preocupación sustantiva compartida entre múltiples disciplinas; más bien, se trata de un cambio fundamental en la perspectiva analítica que requiere una actualización exhaustiva del modo de análisis dominante de las disciplinas (Darian-Smith y McCarty 2017, 2). Por ejemplo, si los estudios sobre la globalización se llevaran a cabo exclusivamente dentro de las disciplinas ya establecidas, esto generaría que dichas aproximaciones fueran parciales, ya que estarían limitadas por los cánones teóricos y metodológicos tradicionales que dan sustento a dichas disciplinas y reflejan el tiempo y el lugar en que estas se originaron (Appelbaum 2013, 546). Así, mientras que los estudios sobre la globalización tienen como punto de partida los Estados nación y son normativamente eurocéntricos en sus narrativas, los estudios globales aspiran a superar estas limitaciones y buscan incorporar múltiples conocimientos sobre aquellos grupos que han sido históricamente excluidos de las narrativas disciplinarias dominantes (Appelbaum 2013, 546).

Si se admite que los estudios globales son una disciplina distinta en el campo de las ciencias sociales y las humanidades, esto se debe a la complejidad misma del objeto de estudio —las tendencias globales con sus múltiples matices y problemas— y a la incapacidad de los paradigmas de las disciplinas actuales para comprender dentro de su marco conceptual las dinámicas y tendencias de lo global. Esta incapacidad paradigmática se deriva tanto de las conceptualizaciones previas del saber, como de la metodología empleada para conocer con precisión lo que creemos saber con cierto grado de certeza. Por lo tanto, se sugiere que el surgimiento y la consolidación de los estudios globales como una nueva disciplina científica reflejan las incapacidades paradigmáticas, tanto ontológica como epistemológica, que hoy en día encierran los modelos teóricos metodológicos dominantes en las disciplinas actuales en el campo de las ciencias sociales.

Esta doble incapacidad paradigmática puede ser una ocasión para explorar otras formas de conceptualizar, no solo la esencia del objeto de estudio al que llamamos *lo global*, sino también las propuestas metodológicas cuya finalidad es encontrar formas más precisas de aprehender este fenómeno y explicarlo de manera rigurosa, congruente y consistente. Por lo tanto, se sugiere que las incapacidades paradigmáticas son una oportunidad para imaginar cómo se pueden ampliar las fuentes de información a nuestro alcance y cómo podemos llevar a cabo otros procedimientos que permitan capturar nueva información, procesarla, analizarla e interpretarla. Bajo este supuesto, es posible reflexionar sobre si hay algo novedoso que pueda ofrecer la ciencia de datos a la disciplina de los estudios globales y si puede enriquecer o complementar los enfoques metodológicos dominantes desde los que tradicionalmente se han analizado la globalización y sus implicaciones.

Al mismo tiempo, estas incapacidades paradigmáticas permiten también plantear la inquietud intelectual sobre hasta qué punto podrían modificarse las bases ontológicas de lo global si empleamos nuevos enfoques basados en la ciencia de datos. En otras palabras, ¿qué sucederá si somos capaces de afinar la búsqueda y la correlación de datos masivos a nivel local, nacional o mundial para abordar problemas de investigación relacionados con las tendencias globales? ¿Cambiarán nuestras perspectivas de lo global o simplemente se “expandirá” nuestro conocimiento de lo que ya sabemos? ¿La ciencia de datos proporcionará nuevos patrones para comprender mejor las tendencias globalizantes y sus implicaciones o representará una nueva forma eurocéntrica del conocimiento científico que disminuya o limite otras voces no occidentales del conocimiento que los estudios globales se han esforzado en retomar? Con la intención de bosquejar algunas respuestas al respecto, en la siguiente sección se analizarán las posibles implicaciones metodológicas de la ciencia de datos para los estudios globales.

3. Las implicaciones teórico-metodológicas de la ciencia de datos para los estudios globales

Hoy en día es posible examinar las actividades del ser humano a escalas inimaginables a partir de las “huellas digitales”, las cuales tienen el potencial de ayudar a comprender mejor las complejidades del comportamiento humano (Lin 2015, 34). El nuevo tipo de información derivado del *big data* ha originado un campo de investigación completamente diferente; a la vez, las ciencias informáticas han generado otras formas de crear y recopilar datos, han desarrollado técnicas analíticas y estadísticas, y han proporcionado formas inéditas de visualizar y presentar la información. En pocas palabras, estas técnicas y fuentes de información tienen el potencial de transformar las metodologías de las ciencias sociales (Foster *et al.* 2016, 1). De hecho, a partir de las contribuciones de la ciencia de datos han surgido aproximaciones que están cambiando los objetos del conocimiento, al mismo tiempo que generan nuevas narrativas sobre la manera en la que comprendemos las interacciones sociales y las relaciones de los seres humanos con su entorno (Boyd y Crawford 2012, 665). Así, es factible suponer que, si somos capaces de mejorar la estructura de la información previa sobre la cual se basan nuestras estimaciones, se podrá reducir las incertidumbres sobre el conocimiento obtenido (Hilbert 2016, 140).

Sin embargo, todo enfoque y toda herramienta metodológica representan una manera parcial de capturar y analizar aquella información que es adecuada a cierto marco conceptual desde el cual es posible analizar tendencias, establecer relaciones, generar conclusiones, construir explicaciones plausibles y así contribuir a nuestros modelos mentales que dan sentido y significado al mundo que nos rodea, favoreciendo la construcción de una cierta “imagen científica del mundo”. En este sentido, la ciencia de datos, como cualquier otra metodología científica, es de crucial importancia para la construcción de la ciencia en su conjunto y de las distintas disciplinas del saber; a la vez, coadyuva a descubrir cuál es la naturaleza de la “imagen científica del mundo”, cuál es la sustancia de la “ley científica” y qué métodos y procedimientos podrían y deberían usarse dentro de la investigación científica (Ochraña 2015, 7-8). Al mismo tiempo, todo enfoque metodológico está embebido en un contexto histórico y social y es moldeado por tradiciones nacionales e influencias transnacionales (Della Porta y Keating 2008, 5). Por lo tanto, el uso masivo de datos y las técnicas de investigación asociadas a su obtención e interpretación no son ajenos a las limitaciones intrínsecas de cualquier método de investigación científica derivadas de los contextos específicos en que surgen.

A pesar de que la ciencia de datos como metodología de investigación social puede ofrecer ciertas contribuciones novedosas al campo de los estudios globales, ya que permite obtener mayores cantidades de información y establecer correlaciones y patrones más robustos con respecto a las dinámicas globalizantes de nuestras sociedades, no está exenta de riesgos. De hecho, cada vez con mayor frecuencia, los algoritmos median los procesos sociales, las transacciones comerciales, las decisiones gubernamentales y la manera en que percibimos, entendemos e interactuamos entre nosotros y con el medio ambiente (Mittelstadt, *et al.*, 2016, 1). En consecuencia, si nuestra comprensión sobre las brechas entre el diseño y el funcionamiento de los algoritmos y sus implicaciones éticas carece de una postura crítica, se generarán graves consecuencias que afectarán tanto a individuos como a grupos y sociedades enteras.

En este sentido, la pregunta fundamental refiere a qué puede aportar específicamente el análisis de datos masivos a la forma tradicional de obtener información con respecto a las tendencias globalizantes de nuestro mundo. Y, en esa medida, de qué manera esta forma de obtener y analizar la información enriquece las aportaciones que los estudios globales llevan a cabo como disciplina científica. En otras palabras, cómo el análisis de datos masivos puede enriquecer el enfoque interdisciplinario, la óptica multicéntrica y la perspectiva multinivel que distinguen el campo de los estudios globales.

En primer lugar, el análisis de datos masivos permite la visualización de tendencias no imaginadas, las cuales surgen separadas de nuestros patrones mentales previos. A diferencia de las metodologías dominantes de las ciencias sociales, en las que se plantean hipótesis a partir de generalizaciones derivadas de ciertas teorías previamente establecidas que permiten orientar la búsqueda de evidencia empírica, los estudios basados en la ciencia de datos se sumergen en un océano de información sin una guía previa de lo que se debe buscar y encontrar. Es más bien una inmensa red que se lanza al mar sin una indagación predirigida. En comparación con las metodologías tradicionales, cuantitativas o cualitativas, las cuales previamente deciden qué datos buscar (Arteaga, Bermúdez y Campos 2004, 15), el análisis masivo de datos procura atraer la mayor cantidad de datos posibles. Así, debido a la posibilidad de procesarlos en cantidad y tiempo extraordinarios, se bosquejan asociaciones entre los datos que, incluso, probablemente no habían sido previamente imaginadas (Attewell y Monaghan 2015, 3). Para ilustrar esta situación en el campo de los estudios globales, es posible situar como ejemplo el fenómeno de la migración. Las causas y motivaciones de la migración mundial han sido explicadas desde múltiples perspectivas. Pero, si podemos seguir los datos de los migrantes a través del consumo de datos que realizan a través de sus teléfonos celulares, ¿qué nuevas posibilidades interpretativas surgirían con

respecto a desplazamiento, estancia, consumo, expectativas y toma de decisiones? ¿Qué nos dicen todos estos datos en su conjunto cuando los contrastamos masivamente? ¿Qué correlaciones pueden surgir? ¿Y qué similitudes y discrepancias podemos identificar entre los diferentes fenómenos migratorios presentes a escala planetaria, por ejemplo, entre los migrantes del Magreb y los de América Central? ¿Qué elementos hay en común? ¿En qué son diferentes?

Una segunda aportación refiere a que los datos masivos permiten identificar patrones que hacen posible analizar mejor los fenómenos locales. El hecho de que tradicionalmente la obtención y el procesamiento de datos sea una labor onerosa ha propiciado que la mayor parte de los datos se presenten a nivel nacional. El estudio de casos ha sido visto como un camino para conocer a profundidad dinámicas en ciertas regiones o localidades; sin embargo, el estudio de casos se considera más una aproximación ilustrativa, que no permite obtener generalizaciones, debido a la imposibilidad de contrastar múltiples casos y múltiples variables (Saldaña 2011, 9-10). Pero ahora con los datos masivos es posible establecer patrones locales y contrastarlos con los nacionales y transnacionales. No hay que olvidar que gran parte de los datos masivos son sociales, es decir, sobre las interacciones de las personas: cómo se comunican, cómo forman relaciones, cómo entran en conflicto y cómo configuran sus futuras interacciones a través de instituciones políticas y económicas (Monroe *et al.* 2015, 71). Con la intención de plantear un ejemplo para la discusión, se puede indagar patrones de consumo cultural de escala global vinculados a contextos locales específicos. Así, es posible explorar cuestiones como: ¿Cuáles son las tendencias de discusiones en redes sociales como Twitter? ¿Cuáles tendencias nacionales y mundiales repercuten en contextos locales? ¿Existen correlaciones profundas? ¿Y, cuando se establecen estas correlaciones, a qué ámbitos sociales pertenecen? ¿Existe simultáneamente influencia en las grandes tendencias mundiales de discusión desde el sur global o solamente se derivan del norte global? ¿Y, si así sucediera, qué patrones se observan que permitan explicar estos fenómenos?

Además, el uso masivo de datos hace posible generar múltiples variables de modelación y supera el principio de parsimonia. Este establece que, dadas las mismas condiciones, la explicación que suele resultar correcta es la más sencilla (Everitt 2012, 278). En cierta medida se comprende que, dadas condicionantes como el tiempo, el costo y el espacio, que tradicionalmente han limitado la cantidad de información que se puede analizar, se adoptaran modelos que tendieran a ser más sencillos y que intencionalmente rehusaran la complejidad. Sin embargo, en la medida en que la revolución tecnológica avanza, la complejidad en el manejo y procesamiento de datos se irá reduciendo de manera significativa. Por lo tanto, será posible acercarse a modelos complejos que procesen múltiples variables a

escala local, nacional y transnacional estableciendo patrones de interrelación más complejos. En este sentido, se puede señalar que existe una amplia literatura que ha centrado su investigación en los movimientos sociales y la lucha por los derechos humanos a escala planetaria. Pero ¿qué puede aportar la ciencia de datos al abordar la apropiación y defensa de estas causas cuando se parte de la perspectiva del ciberactivismo? ¿Qué otra información podemos obtener de aquellos perfiles que llevan a cabo actividades pro defensa de los derechos humanos utilizando redes sociales? ¿Qué otros comportamientos tienen en la red que ilustren mejor la manera en que el ciberactivista relaciona su contexto local con las demandas mundiales? ¿Cómo esos comportamientos generan patrones de interrelación y reacción? ¿Cómo se correlacionan esos factores?

Finalmente, el análisis de datos masivos en tiempo real permite acceder al flujo continuo de información, superando las cadenas de temporalidades prefijadas, lo que representa un camino que posibilita el abandono de linealidades. Tradicionalmente, debido a la practicidad que esto representa, la obtención de información tiene un corte temporal, como las estadísticas de la actividad productiva y los censos de población. Sin embargo, debido a la mayor automatización de la recopilación y el análisis de datos, así como a los algoritmos que pueden extraer e ilustrar patrones a gran escala, es posible pensar en nuevos cortes temporales (Boyd y Crawford 2012, 4). Ahora, el constante flujo de información e incluso la posibilidad de su análisis en tiempo real permiten que la temporalidad de su estudio sea limitada solamente por las pautas establecidas por el investigador. Por ejemplo, ¿cómo el seguimiento del surgimiento y la expansión de los brotes de infección a partir de miles de historias clínicas en distintas partes del mundo pueden ampliar nuestra comprensión sobre cómo lo global conecta con lo local y viceversa, más allá de los eventos clínicos extraordinarios e irregulares como la expansión de las pandemias? ¿Qué variables están relacionadas entre sí en este y otros fenómenos? ¿Qué nos puede aportar para nuestras explicaciones la información georreferenciada derivada de los desplazamientos de enfermos y profesionales de la salud?

Sin embargo, sería un error considerar el uso masivo de datos como la panacea para comprender las dinámicas globalizantes en nuestra sociedad contemporánea. Al respecto, nosotros identificamos al menos cuatro grandes riesgos de un uso no crítico de análisis de datos masivos. En primer lugar, un regreso a una nueva versión del positivismo y la consolidación de una epistemología empirista como forma superior de conocimiento. Aunque la ciencia de datos pretende ser exhaustiva en la recolección de la información y proporcionar una resolución completa del fenómeno a analizar, esto no deja de ser una representación y una muestra conformada por una plataforma tecnológica, por la ontología

de los datos empleada y por el entorno regulatorio en el acceso a los datos; por lo tanto, la obtención de información está sujeta al sesgo de muestreo (Crawford 2013). De hecho, todos los datos proporcionan puntos de vista oligópticos del mundo: puntos de vista desde ciertos puntos de vista, utilizando herramientas particulares (Haraway 1991, 152). Como tales, los datos no son simplemente elementos naturales que se abstraen del mundo de manera neutral y objetiva ni pueden aceptarse al pie de la letra, como verdades absolutas; al contrario, se crean dentro de un conjunto complejo que da forma activa a su constitución (Ribes y Jackson 2013, 149).

Un segundo riesgo es caer en los excesos de una ciencia predictiva que parte del supuesto de que los seres humanos actuamos bajo ciertos patrones que pueden ser anticipados. Estos supuestos sostienen que existen formas de comportamiento humano predictivas que pueden ser descubiertas si se cuenta con algoritmos adecuados, a partir del análisis de las pautas previamente registradas. En el fondo, la premisa que está detrás de esta apreciación es que, entre mayor cantidad de datos sean analizados bajo algoritmos cada vez más precisos, mayor será la predicción que se pueda obtener. Aunque existe un amplio debate entre el libre albedrío y el determinismo estructural (Honderich 2005, 567-570), estas aproximaciones pueden favorecer el resurgimiento de una cierta fascinación por descubrir las leyes del comportamiento humano. En otras palabras, se piensa que en la medida en que los algoritmos logran generar pronósticos cada vez más precisos de los fenómenos naturales, como el clima, podrán predecir de manera más precisa el comportamiento de las masas humanas en contextos predeterminados. Por ejemplo, en el campo de la seguridad mundial se ha sugerido que el análisis predictivo basado en la ciencia de datos “promete asegurar el futuro al anticipar el ‘próximo ataque terrorista’ y aprehender a posibles delincuentes antes de que puedan atacar” (Aradau y Blanke 2016, 374). De esta manera, se podría concluir falazmente que, con el análisis de datos masivos, por fin será posible establecer patrones definitivos que expliquen el comportamiento social en contextos locales, nacionales y globales; aún más, sugerir que estos algoritmos no solamente establecerán lo que hacemos sino lo que debemos hacer. Un riesgo al respecto es que nuestro horizonte de decisiones se verá acortado a partir de patrones de información prestablecidos. Los nuevos análisis pueden presentar la ilusión de descubrir automáticamente ideas sin hacer preguntas pero, como lo ha advertido Kitchin (2014, 5), los algoritmos utilizados ciertamente surgieron y fueron probados científicamente para verificar su validez y veracidad en contextos específicos, los cuales no son necesariamente universales.

Asimismo, la ciencia de datos puede conducir a un falso dilema, entre la causalidad y la correlación como la mejor manera para comprender la realidad

contemporánea (Cowsly y Schroeder 2015, 448). En este sentido, el riesgo es que se pretenda preferir un aspecto de la construcción de explicación sobre el otro, privilegiando tendencias correlacionales que impiden establecer explicaciones basadas en la relación entre causas y consecuencias. Sin embargo, los datos no hablan por sí mismos. La correlación no implica causalidad. El análisis de datos a gran escala puede detectar correlaciones entre una multitud de fuentes de información, pero no establecer si las correlaciones son significativas. Además, si buscamos correlaciones entre variables, se espera que las pruebas de significación estadística encuentren relaciones falsas basadas únicamente en el azar (Shan *et al.* 2015, 39). Los datos deben ser narrados, puestos a trabajar en contextos particulares, hundidos en narrativas que les den forma y significado y movilizados como parte de procesos más amplios de interpretación y creación de significado (Dourish y Gómez 2018, 1). Aunque la ciencia como sistema de enunciados sobre la realidad es axiológicamente neutra, la elección sobre los fines científicos es un problema filosófico y político (Catalán y Catany 1986, 164). Evidentemente, estos fines se reflejan en ciertas construcciones narrativas que no son neutrales. La ciencia de datos, en el fondo, no es más que una narrativa entre otras.

Finalmente, existe el riesgo de la falsificación de imágenes de la realidad, a partir de la reducción de lo real a aquello que se expresa en los flujos masivos de información que en la era de las tecnologías de la información generamos a través de las interacciones digitales o de nuestras pautas de consumo. No se debe olvidar que las dinámicas de exclusión de nuestro mundo ahora también están generando un nuevo tipo de descartados, los descartados digitales cuyas “huellas” de datos en las nubes son limitadas o inexistentes. Si privilegamos el uso masivo de datos, podríamos caer en una nueva forma de invisibilidad. Este riesgo es aún mayor que el analfabetismo digital, ya que se trata de una nueva forma de invisibilidad, exclusión y segregación.

Conclusiones

La era de los datos masivos ya ha llegado. Los datos masivos continuarán creciendo en los próximos años, lo que se traducirá en información más valiosa, como nunca antes se había visto. Los nuevos conocimientos que aporten los grandes conjuntos de datos pueden ser útiles para dar sentido a muchos aspectos de la vida ordinaria, desde la información que proporcionan los datos de los sensores que registran la salud de las personas hasta aquellos generados en las redes sociales. El análisis de este tipo de información es valioso para comprender las dinámicas actuales de la sociedad y las tendencias globales en que dichas dinámicas se encuentran embebidas. De hecho, el *big data* no solo se ha convertido

en una potente herramienta para el análisis de las tendencias modernas, sino que también ha comenzado a afectar a casi todos los dominios y aspectos de nuestra vida y, en última instancia, tiene el potencial de cambiar la forma en que vivimos (Gupta y Rani 2019, 337-338).

Específicamente en el campo de los estudios globales, la posibilidad de obtener y procesar datos masivos podría revolucionar la manera en que comprendemos ciertas tendencias y dinámicas de la globalización, tanto en una dimensión mundial como en las interacciones entre lo global con lo local o regional. Así, es posible esperar que la incorporación de técnicas asociadas a las ciencias de datos genere ciertos análisis cada vez más precisos. Las posibilidades metodológicas que se abren con la ciencia de datos puede ser un paso importante para analizar información que origine nuevas explicaciones sobre los fenómenos mundiales.

Sin embargo, para que las herramientas tecnológicas basadas en *big data* enriquezcan nuestro conocimiento y nuestra comprensión en el campo de los estudios globales, es necesario mantener una actitud crítica, sin olvidar que tanto la elección de los datos como su análisis y las narrativas que se construyen alrededor de los flujos de información no dejan de ser elecciones que no son totalmente objetivas. Los datos, por más vastos que sean, no significan nada hasta que les damos una interpretación, la cual se formula y reformula desde contextos específicos del saber. Así, si queremos enriquecer metodológicamente nuestra comprensión de lo global, es fundamental mantener un rumbo preestablecido que oriente el quehacer de la investigación científica sin menoscabo de la posibilidad de ampliar nuestro horizonte metodológico que pueda derivar en formas distintas de análisis de la información.

Desde nuestra perspectiva, el rumbo que debe orientar las técnicas de análisis de datos masivos se tiene que sustentar en los tres pilares de los estudios globales. En primer lugar, hay que asumir un enfoque interdisciplinario, utilizar una óptica multicéntrica y partir desde una perspectiva multinivel con la intención de reducir el sesgo de los datos a analizar. Segundo, es necesario adoptar una visión histórica que permita establecer secuencias en una perspectiva histórica de larga duración o que facilite la comparación de procesos globalizantes históricos y contemporáneos, con la intención de superar posibles visiones restringidas de la realidad que lleven a pensar que lo real es simplemente el resultado de un mundo de información reciente que se “descubre” a partir del análisis masivo de datos. Tercero, es fundamental tener en cuenta que existe más de una globalización y es preciso no olvidar la mirada del sur global, donde otras pautas de interacción y dinámicas sociales que se desprenden de contextos específicos no pueden ser captados de la misma manera por el rastro que queda reflejado en las “huellas” digitales. En este sentido, las “huellas” de los millones de bytes de información del

sur global no pueden ser idénticas a las de otras sociedades. Si se pierden estas orientaciones que se fundamentan en la especificidad de los estudios globales como un campo del saber científico, el uso de datos masivos no será más que un nuevo canto de sirenas, con un enorme poder seductor, que nos presentará una visión distorsionada de la realidad.

Referencias

1. Appelbaum, Richard P. 2013. "Comments on Jan Nederveen Pieterse's Essay, 'What is Global Studies?'" *Globalizations* 10 (4): 545-550. <https://10.1080/14747731.2013.806727>
2. Aradau, Claudia y Tobías Blanke. 2016. "Politics of Prediction: Security and the Time/Space of Governmentality in the Age of Big Data". *European Journal of Social Theory* 20 (3): 373-391. <https://doi.org/10.1177/1368431016667623>
3. Araya, Agustín A. 2005. "Internet". En *Encyclopedia of Science, Technology, and Ethics*, editado por Carl Mitcham, 1048-1052. Detroit, MI: Macmillan Reference
4. Arteaga, Carlos, Roberto Bermúdez y Guillermo Campos. 2004. *Ciencias sociales e investigación social*. Ciudad de México: Universidad Nacional Autónoma de México.
5. Attewell, Paul y David Monaghan. 2015. *Data Mining for the Social Sciences: An Introduction*. Oakland: University of California Press.
6. Barranco, Ricardo. 2012. "¿Qué es Big Data? Todos formamos parte de ese gran crecimiento de datos". IBM. Consultado el 27 de enero de 2020. <https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/index.html>
7. Boyd, Danah y Kate Crawford. 2012. "Critical Questions for Big Data". *Information, Communication & Society* 15 (5): 662-679. <https://doi.org/10.1080/1369118X.2012.678878>
8. Campbell, Patricia J., Aran MacKinnon y Christy R. Stevens. 2010. *An Introduction to Global Studies*. Malden: Wiley-Blackwell.
9. Catalán Fernández, Albert y M. Catany Escandell. 1986. "Contra el mito de la neutralidad de la ciencia. El papel de la historia". *Enseñanza de las Ciencias: Revista de Investigación y Experiencias Didácticas* 4 (2): 163-166.
10. Chitra, Kandan y Balakrishnan Subashini. 2013. "Data Mining Techniques and its Applications in Banking Sector". *International Journal of Emerging Technology and Advanced Engineering* 3(8): 220-222.
11. Clare, Susan E. y Pamela L. Shaw. 2016. "Big Data" for Breast Cancer: Where to Look and what You Will Find". *NPI Breast Cancer* 2: 1-5. <https://doi.org/10.1038/npjbcancer.2016.31>
12. Cows, Josh y Ralph Schroeder. 2015. "Causation, Correlation, and Big Data in Social Science Research". *Policy & Internet* 7 (4): 447-472. <https://doi.org/10.1002/poi3.100>
13. Cox, Michael y David Ellsworth. 1997. "Application-Controlled Demand Paging for Out-of-Core Visualization". <https://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>
14. Crawford, Kate. 2013. "The Hidden Biases of Big Data". *Harvard Business Review* 1: 2013. <https://hbr.org/2013/04/the-hidden-biases-in-big-data>
15. Darian-Smith, Eve y Philip C. McCarty. 2017. *The Global Turn: Theories, Research Designs, and Methods for Global Studies*. Berkeley, CA: University of California Press.

16. Davenport, Thomas H. y Dhanurjay Patil. 2012. "Data Scientist: The Sexiest Job of the 21st Century". *Harvard Business Review* 90 (10): 5-7.
17. Della Porta, Donatella y Michael Keating. 2008. Introducción a *Approaches and Methodologies in the Social Sciences*, editado por Donatella della Porta y Michael Keating, 1-15. Cambridge: Cambridge University Press.
18. Dourish, Paul y Edgar Gómez Cruz. 2018. "Datafication and Data Fiction: Narrating Data and Narrating with Data". *Big Data & Society* 5 (2): 1-10. <https://doi.org/10.1177/2053951718784083>
19. Drucker, Peter F. 1993. *Post-Capitalist Society*. Oxford: Butterworth-Heinemann.
20. Everitt, Brian S. 2002. "Parsimony Principle". En *Cambridge Dictionary of Statistics*, editado por Brian S. Everitt, 272. Cambridge: Cambridge University Press.
21. Few, Stephen. 2013. "Data Visualization for Human Perception". En *The Encyclopedia of Human-Computer Interaction*, editado por Constantine Stephanidis, capítulo 35. Aarhus, Dinamarca: The Interaction Design Foundation.
22. Foster, Ian, Rayid Ghani, Ron S. Jarmin, Frauke Kreuter y Julia Lane. 2016. *Big Data and Social Science: A Practical Guide to Methods and Tools*. Nueva York: Chapman & Hall.
23. Grable, John E. y Angela C. Lyons. 2018. "An Introduction to Big Data". *Journal of Financial Service Professionals* 72 (5): 17-20.
24. Graham, Williams. 2011. *Data Mining with Rattle and R. The Art of Excavating Data for Knowledge Discovery*. Nueva York: Springer Science Business Media.
25. Gupta, Deepak y Rinkle Rani. 2019. "A Study of Big Data Evolution and Research Challenges". *Journal of Information Science* 45 (3): 322-340. <https://doi.org/10.1177/0165551518789880>
26. Han, Jiawei, Michelline Kamber y Jian Pei. 2011. *Data Mining: Concepts and Techniques*. Nueva York: Springer.
27. Hand, David J., Heikki Mannila y Padhraic Smyth. 2001. *Principles of Data Mining*. Cambridge: A Bradford Book-The MIT Press.
28. Haraway, Donna. 1991. *Simians, Cyborgs and Women: The Reinvention of Nature*. Nueva York: Routledge.
29. Hilbert, Martin. 2016. "Big Data for Development: A Review of Promises and Challenges". *Development Policy Review* 34 (1): 135-174. <https://doi.org/10.1111/dpr.12142>
30. Hilbert, Martin y Priscila López. 2011. "The World's Technological Capacity to Store, Communicate, and Compute Information". *Science* 332 (6025): 60-65. <https://doi.org/10.1126/science.1200970>
31. Honderich, Ted. 2005. "Determinism". En *New Dictionary of the History of Ideas*, editado por Maryanne Cline Horowitz, 567-570. Detroit, MI: Charles Scribner's Sons.
32. Juergensmeyer, Mark. 2012. "Global Studies". En *Encyclopedia of Global Studies*, editado por Mark Juergensmeyer y Helmut Anheier, 849-861. Thousand Oaks, CA: Sage.
33. Juergensmeyer, Mark. 2013. Introducción a *Thinking Globally: A Global Studies Reader*, editado por Mark Juergensmeyer, 3-9. Berkeley, CA: University of California Press.
34. 34. Kitchin, Rob. 2014. "Big Data, New Epistemologies and Paradigm Shifts". *Big Data & Society* 1 (1): 1-12. <https://doi.org/10.1177%2F2053951714528481>

35. Klösgen, Willi y Jan M. Zytkow. 2002. "The Knowledge Discovery Process". En *Handbook of Data Mining and Knowledge Discovery*, editado por Willi Klösgen, Jan M. Zytkow y Jan M. Zytkow, 10-21. Oxford: Oxford University Press.
36. Lafrate, Fernando. 2018. *Artificial Intelligence and Big Data: The Birth of a New Intelligence*. Hoboken, NJ: Wiley-ISTE.
37. Larose, Daniel T. y Chantal D. Larose. 2014. *Discovering Knowledge in Data: An Introduction to Data Mining*. Hoboken, NJ: John Wiley & Sons.
38. Lin, Jimmy. 2015. "On Building Better Mousetraps and Understanding the Human Condition: Reflections on Big Data in the Social Sciences". *The Annals of the American Academy of Political and Social Science* 659 (1): 33-47. <https://doi.org/10.1177/0002716215569174>
39. Mathivanan, Sandeepkumar y Prabhu Jayagopal. 2019. "A Big Data Virtualization Role in Agriculture: A Comprehensive Review". *Walailak Journal of Science & Technology* 16 (2): 55-70.
40. Mayer-Schönberger, Viktor y Kenneth Cukier. 2013. *Big Data: A Revolution that Will Transform how We Live, Work, and Think*. Londres: John Murray Publisher.
41. Mieczkowski, Yaneck. 2013. *Eisenhower's Sputnik Moment: The Race for Space and World Prestige*. Ithaca, New York: Cornell University Press.
42. Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter y Luciano Floridi. 2016. "The Ethics of Algorithms: Mapping the Debate". *Big Data & Society* 3 (1): 1-21. <https://doi.org/10.1177/2053951716679679>
43. Monroe, Burt L., Jennifer Pan Ha, Margaret E. Roberts, Maya Sen y Betsy Sinclair. 2015. "No! Formal Theory, Causal Inference, and Big Data Are not Contradictory Trends in Political Science". *PS: Political Science & Politics* 48 (1): 71-74. <http://j.mp/20Ft5VN>
44. Naciones Unidas. 2014. "Data Revolution Report: A World That Counts". Naciones Unidas. Consultado el 27 de enero de 2020. <http://www.undatarevolution.org/report/>
45. Nederveen Pieterse, Jan. 2013. "What is Global Studies?". *Globalizations* 10 (4): 499-514. <https://doi.org/10.1080/14747731.2013.806746>
46. Nunan, Daniel y María Laura di Domenico. 2013. "Market Research and the Ethics of Big Data". *International Journal of Market Research* 55 (4): 505-520. <https://doi.org/10.2501/IJMR-2013-015>
47. Ochrana, Frantisek. 2015. *Methodology of Social Sciences*. Praga: Karolinum Press.
48. Ribes, David y Steven J. Jackson. 2013. "Data Bite Man: The Work of Sustaining Long-Term Study". En *Raw Data is an Oxymoron*, editado por Lisa Gitelman, 147-166. Cambridge, MA: The MIT Press.
49. Riquelme, José C., Roberto Ruiz y Karina Gilbert. 2006. "Minería de datos: conceptos y tendencias". *Inteligencia Artificial* 10 (29): 11-18.
50. Ryan, Johnny. 2009. *A History of the Internet and the Digital Future*. Londres: Reaktion Books Limited.
51. Saldaña, Johnny. 2011. *Fundamentals of Qualitative Research*. Nueva York: Oxford University Press.
52. Saouabi, Mohamed y Ezzati Abdellah. 2019. "Proposition of an Employability Prediction System Using Data Mining Techniques in a Big Data Environment". *International Journal of Mathematics & Computer Science* 14 (2): 411-424.
53. Shan, Dhavan V., Joseph N. Cappela, Neuman Russel y Jimmy Li. 2015. "On Building Better Mousetraps and Understanding the Human Condition: Reflections on Big Data in the Social Sciences". *Annals* 659: 33-47. <https://doi.org/10.1177/0002716215569174>

54. Simpson, Dana. 2016. *The Use of Big Data: Benefits, Risks, and Differential Pricing Issues*. Nueva York: Nova Science Publisher.
55. Van der Aalst, Wil. 2014. "Data Scientist: The Engineer of the Future". En *Enterprise Interoperability IV*, editado por Kai Mertins, Frédérick Bénaben, Raúl Poler y Jean-Paul Bourrières, 13-28. Nueva York: Springer.
56. Ware, Colin. 2012. *Information Visualization: Perception for Design*. Ámsterdam: Elsevier.



Daniel Lemus-Delgado es profesor investigador de la Escuela de Ciencia Sociales y Gobierno del Tecnológico de Monterrey. Doctor en Relaciones Internacionales Transpacíficas, por la Universidad de Colima. Es profesor titular de las materias Teoría de las Relaciones Internacionales, Análisis de Política Exterior y Escenario Asia Pacífico. Ha sido *Visiting Scholar* en la Universidad de Fudan, Shanghai, en China. Entre sus líneas de investigación se encuentran Teorías de Relaciones Internacionales y Política Exterior. Es investigador nivel II del Sistema Nacional de Investigadores del Consejo Nacional de Ciencia y Tecnología. ✉ dlemus@tec.mx

Ricardo Pérez Navarro es doctor en Ciencias Financieras por el Tecnológico de Monterrey y economista y Maestro en Economía por la Universidad de Guadalajara. Investigador del Sistema Nacional de México y profesor en el Tecnológico de Monterrey campus Guadalajara en asignaturas a nivel profesional relacionadas finanzas y riesgos, así como en postgrado la EGADE Business School en las asignaturas de economía y estadística. Actualmente sus líneas de investigación están relacionadas con el área de riesgos financieros y la ciencia de datos. Ha publicado artículos académicos en el área de finanzas, en revistas como *Network Industries Quarterly*, *Revista EconoQuantum* y *Revista Estocástica* entre otras y de igual manera cuenta con distintos capítulos en libros de áreas similares. ✉ ricardo.perez@tec.mx