

**BIG DATA, PROBLEMAS EPISTEMOLÓGICOS Y  
CONTEXTUALISMO EPISTÉMICO**

**BIG DATA, EPISTEMOLOGICAL PROBLEMS AND  
EPISTEMIC CONTEXTUALISM**

**BIG DATA, PROBLEMAS EPISTEMOLÓGICOS E  
CONTEXTUALISMO EPISTÊMICO**

**José Carlos Chahuara Quispe\***

Universidad Nacional Mayor de San Marcos  
jose.chahuara@unmsm.edu.pe  
ORCID: 0000-0002-8003-8804

Recibido: 14/10/22

Aprobado: 22/11/22

---

\* José Carlos Chahuara Quispe es estudiante de la Maestría en Filosofía con mención en Epistemología de la Universidad Nacional Mayor de San Marcos. MBA e Ingeniero Electrónico de profesión se encuentra realizando una tesis sobre la evidencia empírica en la Ciencia de Datos a la que el presente artículo se encuentra relacionado.

## Resumen

La *big data*, las computadoras y la ciencia de datos se han convertido en instrumentos clave para la generación de conocimiento. En este contexto se hace interesante entender cómo estas metodologías influyen en lo que los usuarios de distintos ámbitos (científico, económico, financiero, etc.) pueden afirmar respecto a qué es lo que conocen y lo que no. El presente artículo describe los problemas epistemológicos derivados del uso de *big data*; se propone además plantear en qué medida esta nueva realidad tecnológica puede ser una fuente de contextualismo epistémico, y cómo, utilizando las reglas propuestas por Lewis (1996), se puede encontrar un punto de entendimiento entre quienes ven el mundo como un enorme conjunto de datos que contienen la respuesta a los problemas, y quienes creen que solo el método científico y las pruebas de hipótesis pueden lograr el conocimiento más certero sobre los fenómenos del mundo.

**Palabras claves:** epistemología, big data, ciencia de datos, contextualismo.

## Abstract

Big data, and data science have become key instruments for the generation of knowledge. In this context, it is interesting to understand how these methodologies influence what users from different fields (scientific, economic, financial, etc.) can say about what they know about the world and what they do not know. This article describes the epistemological problems derived from the use of big data, it also proposes to consider to what extent this new technological reality can be a source of epistemic contextualism, and how, using the rules proposed in by Lewis (1996), can find a point of understanding between those who see the world as a huge set of data containing the answer to problems, and those who believe that only the scientific method and hypothesis testing can achieve the most accurate knowledge about the world's phenomena.

**Keywords:** epistemology, big data, data science, contextualism.

## Resumo

O big data e a ciência de dados tornaram-se instrumentos fundamentais para a geração de conhecimento. Nesse contexto, é interessante entender como essas metodologias influenciam o que usuários de diferentes áreas (científicas, econômicas, financeiras, etc.) podem dizer sobre o que sabem e o que não sabem. Este artigo descreve os problemas epistemológicos derivados do uso de big data, também se propõe a considerar em que medida essa nova realidade tecnológica pode ser uma fonte de contextualismo epistémico, e como, usando as regras propostas por

Lewis (1996), pode encontrar um ponto de entendimento entre aqueles que vêem o mundo como um enorme conjunto de dados contendo a resposta para os problemas, e aqueles que acreditam que somente o método científico e o teste de hipóteses podem alcançar o conhecimento mais preciso sobre os fenômenos do mundo.

**Palavras-chaves:** epistemologia, big data, ciência de dados, contextualismo.

## Introducción

El avance tecnológico de la informática y la computación ha cambiado el mundo en el que vivimos en sus diversos ámbitos: social, económico, financiero y científico. En todos ellos, se ha producido una revolución en cuanto a la capacidad para registrar y procesar abundantes cantidades de datos, y a la complejidad de los métodos y algoritmos diseñados para solucionar problemas prácticos, así como para hacer y compartir investigación científica.

La ciencia de datos y la tecnología del *big data* hacen parte de esa revolución tecnológica, y ha encontrado campos de aplicación en diversos sectores de la actividad productiva humana actual como el científico, industrial, financiero, político y de entretenimiento bajo la promesa de obtener conocimiento de forma efectiva y eficiente a partir de *insights* desde el conjunto de datos para la toma de decisiones más acertadas en cuanto al campo empresarial, científico, político, deportivo, entre otros. De esta manera, la *big data* y la ciencia de datos se han convertido en un elemento conjunto que ha creado un nuevo contexto social, cambiando, durante las últimas dos décadas, nuestra existencia y nuestra relación con el mundo, y generando además gran interés entre los filósofos tanto desde los aspectos tecnológico, ético, así como epistemológico.

## Datos, información y conocimiento

Sociedad de la información o sociedad del conocimiento, son términos que se usan para resaltar la importancia, y el impacto

que la revolución impulsada por las tecnologías de la información tienen en el mundo y las sociedades actuales. Sin embargo, el término información se suele confundir con otros términos como dato o conocimiento. Definir qué es información no es una tarea sencilla. En muchos ámbitos, información suele usarse como un término que engloba por sí mismo dato y conocimiento, y además se usa como un concepto que describe la conexión entre los datos sin procesar (*raw data*) y el conocimiento que se alcanza a partir de ellos, luego de realizar algún tipo de análisis (Davenport & Prusak, 1997).

Peter Drucker (1988), definió la información como un conjunto de datos provistos de propósito y relevancia, dando al ser humano el rol de ser quien provee a los datos de esas características, vinculando así a los datos con la inteligencia humana como un requisito para que se conviertan en información. De esta manera, la persona es el elemento activo que es capaz de transformar los datos en información. La información, a diferencia de los puros datos, requiere que se incluyan ciertas características narrativas que permitan análisis y discusión, por ejemplo, en una instancia simple como el índice de variación del precio anual de un producto, siempre podrá ser posible decir algo respecto al puro dato como que tan alto es en relación a lo estimado al inicio del año, o sobre cómo se afecta el poder adquisitivo de la población.

Un paso más allá, el conocimiento también es información, pero información con el mayor de los valores en términos utilitarios, ya que alguien tiene que poner contexto, significado, interpretaciones particulares y argumentación en la información; es decir, alguien tiene que haber reflejado y adicionado su propia sabiduría en la información para hacerla conocimiento y para transmitirla como tal, retomando el ejemplo sobre el índice de precios, el conocimiento se expresaría, por ejemplo, en un argumento a favor o en contra de las políticas económicas de un gobierno. Es claro entonces que la cantidad de participación

del ser humano se incrementa conforme se vaya avanzando en la línea de dato-información-conocimiento.

La importancia de la consideración del factor humano en la discusión respecto a la información, se debe a que la fascinación por la tecnología ha hecho que se olvide el propósito clave de la información, que es transmitir conocimiento a las personas. Todas las computadoras en el mundo no tendrían ninguna utilidad si los usuarios no estuviesen interesados en producir conocimiento a partir de la información generada al utilizarlas. Todos los sistemas de información en los que invierten las empresas no adicionarán ningún tipo valor a su desempeño si los empleados no compartiesen la información que generan en sus funciones con sus colegas de las distintas áreas de la organización. La recolección de datos, la organización de información y la generación de conocimiento son actividades humanas de gran importancia, y el ser humano no será capaz de definirla, entenderla y gestionarla correctamente a menos que la persona tenga el rol primario en la discusión (Davenport & Prusak, 1997). Las computadoras son entonces muy útiles para manejar los datos, no tan útiles para manejar información y aún menos útiles para manejar conocimiento.

## **Big data y ciencia de datos**

El avance explosivo de las capacidades de procesamiento de datos ha originado la aparición de nuevos conceptos y métodos que han catalizado el desarrollo de la informática. Es así que desde inicios del siglo XXI dos de estos conceptos, la ciencia de datos y la *big data*, han encontrado campos de aplicación en diversos sectores de la actividad humana, creando las condiciones para procesar y analizar cantidades ingentes de datos, información, y obtener conocimiento de aplicación práctica.

Mientras es posible entender la *big data* como la existencia y posibilidad de registro, almacenamiento y organización de grandes cantidades de datos, existe una discusión aún abierta

respecto a qué es la ciencia de datos. Una forma de aclarar el concepto, es a través del diagrama de Venn propuesto por Conway (2010) (Figura 1)



Figura 1. Diagrama de Venn de Conway (2010)

El diagrama muestra que la ciencia de datos se encuentra en la intersección de distintas habilidades, como son: a) la programación, es decir el dominio técnico de los sistemas computacionales y sus lenguajes, b) el conocimiento de ciencias formales como las matemáticas y la estadística, y c) la experiencia relevante en investigación científica que permita entender el procedimiento que sigue el descubrimiento científico en cualquiera de sus ramas de aplicación.

A pesar de los esfuerzos de Conway y otros por definir la ciencia de datos, quedan aún muchas dudas y controversias que han impedido alcanzar una definición estandarizada. Sin embargo, es en las tres etapas de procesamiento que sigue la ciencia de datos y que se relaciona con la *big data* donde parece haber consenso: a) recolección de datos, que es la captura y re-

gistro de una gran cantidad de información, en muchos casos no estructurada, a partir de fuentes diversas; b) análisis de datos, que involucra el uso de herramientas como algoritmos, *machine learning* y correlaciones estadísticas; y c) el uso de los datos como una forma de inferencia predictiva (Van Der Sloot, Broeders, & Schrijvers, 2016) de una forma similar a como se usan las teoría científica.

La ciencia de datos sería entonces la ciencia del *big data*, pues a través de su tecnología y herramientas, los usuarios logran procesar el gran volumen de datos disponible, e interpretar la información que contienen para generar conocimiento. Es necesario notar que aunque la definición de ciencia de datos que propuso Conway (2010) refiere a conocimientos y habilidades, estos no tienen que ser teóricos a profundidad, sino que en la mayor parte de casos son conocimiento práctico de las etapas del proceso a seguir y de las condiciones a cumplir en términos de parámetros predefinidos, es decir, no es inusual que un usuario de ciencia de datos y *big data* no sea capaz de explicar el funcionamiento de los algoritmos, pero que si pueda decir siguiendo los parámetros definidos, si lo que ha obtenido al usar ciencia de datos es o no confiable, e interpretar los resultados.

## **Impacto de *big data* y la ciencia de datos en la sociedad**

Los datos, y los sistemas informáticos que hacen uso de esos datos, se encuentran en el ámbito social (Lee & Cook, 2020), a tal punto que se puede afirmar que la ciencia de datos y la *big data* han creado las bases de la nueva sociedad de la información en la que vivimos. Tanto personas como máquinas están conectadas a internet y generan gran cantidad de información a partir de su interacción (Mayer-Schönberger & Cukier, 2013). El impacto que ha generado esta revolución tecnológica es enorme. Por ejemplo, en el campo empresarial, muchas compañías están haciendo uso de las tecnologías de ciencia de datos y *big*

*data* para encontrar patrones y correlaciones estadísticas que les permitan optimizar sus servicios, atender de mejor manera a sus clientes, e incluso personalizar los contenidos que les ofrecen. Incluso los gobiernos están aplicando la ciencia de datos en distinto nivel de profundidad para mejorar su actuación en los ámbitos de la seguridad, la justicia, la salud, la infraestructura y el transporte (Sloot, Broeders, & Schrijvers, 2016).

La *big data* florece en la sociedad gracias a cuatro puntos principales (Succi & Coveney, 2019): i) el explosivo crecimiento en la producción, adquisición y procesamiento de datos, ii) el análisis de patrones en grandes bases de datos utilizando algoritmos que hacen más rápido el modelar el comportamiento de las variables que hacerlo utilizando alguna teoría, iii) se puede aplicar a cualquier disciplina, incluso a aquellas que por su naturaleza compleja no son adecuadas para el tratamiento matemático usual (e.g. biología o el clima), iv) su aplicación inmediata y sencilla en ramas como la política y los negocios a través de análisis de opiniones y sentimientos que abre la puerta a nuevos dominios de interés para el estudio de fenómenos sociales.

En cuanto al impacto en el conocimiento, que es el activo más importante de la sociedad actual, el simple hecho de que se pueda contar con la capacidad de obtener y procesar cada vez más datos, podría significar claramente que como sociedad se logra saber o conocer más. Por ejemplo, cuando una persona utiliza una tarjeta de crédito robada para hacer compras, la adquisición de datos y el monitoreo de las transacciones a través del análisis de la información permite identificar y revelar el fraude instantáneamente creando un conocimiento práctico que permitirá tomar decisiones a todos los afectados por la situación fraudulenta. Este conocimiento práctico es producto de una generalización estadística acerca de una persona, una tarjeta de crédito extraviada o robada y una forma de comportamiento, que permite la inferencia de comportamientos similares de otras personas más adelante. En este sentido, la *big*



*data* y la ciencia de datos se pueden asociar a un concepto de valor social, que se basa en la creación de bienestar en áreas de interés comunitario como son la educación, la salud y la seguridad pública. Por ejemplo para mejorar y hacer más transparente la información pública, y para hacer que los ciudadanos participen más activamente del escrutinio de esta información para prevenir y evitar el fraude y el crimen, mejorar la seguridad nacional, y mejorar la calidad de vida de sus ciudadanos con mejor educación y sistemas de salud (Günther, Rezazade Mehrizi, Huysman, & Feldberg, 2017).

### **Impacto de big data y la ciencia de datos en la comunidad científica**

Existen ramas de la ciencia actual que hacen uso intensivo de datos, por ejemplo la astronomía, la genética, la física de partículas y las ciencias del océano. Por ejemplo, en astronomía, hay un número infinito de cuerpos celestiales cuyo comportamiento es investigado a partir del estudio de sus espectro electromagnéticos, y existen instrumentos que registran y almacenan esa abundante información (Frické, 2015). Es importante mencionar, que a pesar de ser una herramienta de uso cada vez más frecuente, la *big data* ha sido criticada por investigadores del campo científico como Succi & Coveney (2019) a partir del estudio de sistemas complejos. Gran parte de las áreas de investigación científica hacen parte de lo que se conocen como sistemas complejos, en cuatro puntos clave que dan pie a una serie de discusiones respecto a los problemas epistemológicos de la *big data*: i) los sistemas complejos son fuertemente correlacionados, por lo que generalmente no obedecen las estadísticas Gaussianas, ii) ningún conjunto de datos es lo suficientemente grande para sistemas con gran sensibilidad a los errores de datos, iii) la correlación no implica causalidad, el lazo entre ambas se debilita conforme se incrementa el volumen de datos, iv) en un mundo con capacidad finita, tener muchos datos es tan bueno como no tener ningún dato, de esta manera se trata de

desmitificar el poder de la *big data* para su aplicación en problemas científicos reales.

En la línea de las críticas negativas al uso de *big data* en la ciencia, podemos considerar el uso de la Sloan Digital Sky Survey, que es un listado de todos los cuerpos celestes que los astrónomos van descubriendo; asimismo, en la Ocean Observatories Initiative, que almacena observaciones del comportamiento marino a nivel mundial. En ambos casos, como listados, las observaciones son importantes, pero lo que hacen en la práctica es solo recoger y registrar datos y cada vez más datos. No explican por sí mismas alguna teoría, ni resuelven algún problema científico. Es por ello que muchos detractores del uso de la *big data* consideran que en general, en muchos de los campos científicos, esta tecnología provee la posibilidad de crear grandes listas, catálogos y clasificaciones (Frické, 2015) pero no ciencia. Esto es muy válido, en la medida en que lo que hace interesante verificar si realmente se pueden hacer descubrimientos científicos, o si se puede aprender algo nuevo a partir de la gran cantidad de datos, es el trabajo deductivo que puede realizarse a partir de la información, independientemente del volumen de datos, solo este análisis hace posible generar a partir de los datos nuevas teorías, explicaciones o predicciones.

Estas críticas válidas son, sin embargo, incompletas, pues como se ha expuesto anteriormente la *big data* es un fenómeno que no se caracteriza solo por el volumen de información generado o por la única posibilidad de registro de información, sino también por las técnicas y algoritmos de la ciencia de datos que permiten el procesamiento de esa información para facilitar el análisis y la posibilidad de crear inferencias y obtener conocimiento ya sea de tipo científico o no. Además, la cada vez mayor digitalización de la información, el crecimiento en la generación de datos, y las posibilidades de comunicación casi ilimitadas entre profesionales alrededor del mundo, facilita la posibilidad de que los investigadores o usuarios de *big data* y ciencia de datos trabajen juntos en equipos multidisciplinarios para en-

frentar de una mejor manera los desafíos técnicos, éticos y legales que la investigación con estas nuevas tecnologías requiere (Favaretto, de Clercq, Schneble, & Elger, 2020).

## **Problemas epistemológicos del uso de big data**

Como sucede con muchas revoluciones tecnológicas, la importancia de la *big data* y la ciencia de datos se ha extendido más allá del ámbito científico que le dio origen, generando discusiones de segundo orden de tipo filosófico y epistemológico (Succi & Coveney, 2019). Es así que la epistemología de la ciencia de datos se ha convertido en un tema de interés en la comunidad científica con interés filosófico y en la comunidad filosófica desde hace algo más de una década, permitiendo que los investigadores abran diversos tipos de aproximaciones para discusión y estudio, entre ellas:

- El posible fin del método científico a partir del uso de la ciencia de datos, considerando que de acuerdo al uso actual que se tiene de estas técnicas y de *big data*, casi cualquier conclusión puede inferirse acertadamente detectando patrones en las enormes bases de datos que tienen a su disposición los usuarios de la ciencia de datos. Llevada a un extremo, esta idea es semejante a creer que con suficientes datos, los números hablarán por sí mismos, la correlación reemplazará a la causalidad, y la ciencia podrá avanzar incluso sin modelos coherentes o teorías unificadas (Succi & Coveney, 2019); o desde otro punto de vista, considerando que en la actual era de algoritmos sofisticados e información masiva, las teorías, hipótesis, prueba de hipótesis, discusiones respecto a si determinado experimento refuta o confirma una teoría se han vuelto innecesarias (Mazzocchi, 2015).
- La generación de valor económico a partir del uso de la *big data* en las organizaciones empresariales (Günther et al., 2017), ya que al implementar esta tecnología, las

empresas pueden incrementar sus ingresos a partir de aplicaciones como comercio electrónico, gobernanza electrónica, seguridad de la información, toma de decisiones, entre otras.

- El impacto de la ciencia de datos en la sociedad actual en forma de algoritmos, redes sociales, comunicación móvil, e inteligencia artificial que no se encuentran de forma aislada, sino insertos en la sociedad actual (Lee & Cook, 2020), o en la forma en la que estos nuevos métodos afectan las bases institucionales de la sociedad y la política científica, que como actividad social se rige por normas y convenciones que se están modificando desde la aparición de la ciencia de datos (Ahonen, 2015).
- la influencia de la ciencia de datos en el proceso de toma de decisiones a partir de una definición de lo que el entendimiento de una situación puede significar para un tomador de decisiones, así como lo útil o no útil que puede resultar tener gran abundancia de información antes de elegir un mejor rumbo de acción (Hirschheim, 2021).
- La poca transparencia que hay al interior de los procedimientos que utilizan *big data* tanto desde la disponibilidad de los detalles de los datos, como de su utilidad y manejo (Obar, 2020), como del uso que hacen las personas de herramientas tecnológicas que usan *big data* y ciencia de datos; por ejemplo, el entendimiento de cómo se hace la recolección, limpieza, almacenamiento y presentación de la información en la app Twitter para sus distintos usuarios, que al igual que Facebook son entornos tecnológicos digitales que son inestables y ambiguos por el propio uso que hacen los miembros de estas redes sociales (Driscoll & Walker, 2014).
- Enfoques que sugieren que los estudios y las conclusiones obtenidas con *big data* son insuficientes y requieren una aproximación colaborativa en conjunto con los métodos

científicos tradicionales de tal forma que puedan hacerse análisis, descripción, predicción y control de unidades individuales de estudio (e.g. una persona, una clínica, un hospital, un sistema de salud, etc.), en contraposición con *big data*, que suele hacer uso de una gran cantidad de datos recolectados de un gran grupo de individuos con el objetivo de hacer una descripción y predicción del comportamiento de otro grupo de individuos, no necesariamente aquellos que han provisto los datos (Hekler et al., 2019).

- Definiciones de qué es *big data* en los campos científicos en los que los investigadores están haciendo uso de esta tecnología, como la psicología y la sociología (Favaretto et al., 2020) o como distinción de acuerdo al estado de uso en que se encuentra la información, como datos en principio cuando son registrados, y datos en la práctica cuando son utilizados (Jones, 2019); lo que refleja claramente la dificultad que existe entre los investigadores para dar una definición de lo que es *big data*.

## **Ciencia de datos, *big data* y creencia justificada**

Una creencia racional debe estar justificada por la evidencia disponible para un agente (van Benthem & Pacuit, 2011). Los datos por sí mismos contienen la evidencia, pero no son por sí mismos la evidencia, requieren de la interpretación de las personas, y la existencia de un problema al cual el científico, por ejemplo, quiera dar solución para que los datos conformen una evidencia. Es decir, solo podría haber evidencia cuando se quiere respaldar una determinada posición (proposición) argumentativa respecto a una posible hipótesis, esta proposición es la que da forma a la creencia justificada.

En el caso de la *big data* y las ciencias de datos, esta interpretación puede abrir espacio a una discusión respecto al contextualismo epistémico, que Lewis (1996) propone como una situación en la que “S sabe que P si y solo si la evidencia elimina

toda posibilidad en la cual no-P ... excepto para aquellas posibilidades que están siendo ignoradas apropiadamente”. La discusión se abre paso al considerar que el ignorar posibilidades es un asunto pragmático y contextual, es decir, que depende de la persona que considera que sabe algo. Más aún, si se parte de la idea antigua de que la justificación es la marca que distingue el conocimiento de una mera opinión (incluso verdadera opinión), entonces bien se puede concluir que las adscripciones del conocimiento son dependientes del contexto porque los estándares para una justificación adecuada son dependientes del contexto (Lewis, 1996), y esto debería ser también aplicable al entorno del conocimiento que se genera a partir de la *big data* y las ciencias de datos, considerando que lo distintivo de la ciencia es que se preocupa por la evidencia y está dispuesta a modificar sus teorías en función de la ciencia (Sanjuán, 2021).

### **Aproximación desde el contextualismo epistémico**

El contextualismo epistémico refiere que lo que es expresado como una atribución de conocimiento —por ejemplo el decir que S sabe que ‘p’— depende parcialmente del contexto en el que se encuentra aquel que de quien se dice que sabe algo (Rysiew, 2021). Así, las condiciones de verdad, falsedad y el propio significado de oraciones que involucran términos como “conocimiento” o “saber”, es decir oraciones de la forma “S sabe que p”, puede variar con el contexto del emisor de la oración. Más precisamente, los estándares que alguien debe cumplir para que pueda decir con certeza que “sabe” algo, varían dependiendo del contexto.

Baumann (2016) distinguió 5 parámetros epistémicos que fundamentan la existencia de un contexto: los estándares de evidencia, la credibilidad, los grados de creencia, la posición epistémica, y las alternativas normativas. Estos parámetros permiten describir y explicar las variaciones entre los diferentes estándares de conocimiento, y demostrarían la existencia de

esos distintos estándares. Así mismo, el contextualismo epistémico puede ser visto como una tesis semántica: que concierne a las “condiciones de verdad” de las oraciones que contienen algo referido al “conocimiento” y/o a las proposiciones expresadas verbalmente en ese sentido. Es así que el contextualismo epistémico ha logrado retomar las antiguas preguntas respecto a la posibilidad, la naturaleza y lo relativo del conocimiento humano (Ichikawa, 2017)

En particular, Lewis (1996), defendió la existencia del conocimiento frente al argumento escéptico, que niega tal existencia a partir de la creencia de que solo puede haber conocimiento en cuanto este sea infalible. Para lograrlo, recurre a la idea de que el conocimiento es dependiente del contexto, en relación con la evidencia que se tenga y en la capacidad de esa evidencia de eliminar las posibilidades del no-conocimiento de algo y la propuso como una situación en la que “S sabe que P si y solo si la evidencia elimina toda posibilidad en la cual no-P ... excepto para aquellas posibilidades que están siendo ignoradas apropiadamente”. La discusión se abre paso entonces al considerar que el ignorar posibilidades es un asunto pragmático y contextual, es decir, que depende de la persona que considera que sabe algo, o de la valoración que da a la evidencia que posee. Más aún, si se parte de la idea antigua de que la justificación es la marca que distingue el conocimiento de una mera opinión (incluso verdadera opinión), entonces bien se puede concluir que las adscripciones del conocimiento son dependientes del contexto porque los estándares para una justificación adecuada son dependientes del contexto (Lewis, 1996). Y esto debería ser también aplicable al entorno de la *big data* y las ciencias de datos, considerando la diversidad de técnicas y metodologías usadas para generar conocimiento ya sea de tipo científico o del tipo práctico que se usa en aplicaciones empresariales.

Ahora bien, Lewis propuso reglas que permiten determinar cuáles de las múltiples posibilidades que explican el conoci-

miento de algo podrían no ser apropiadamente ignoradas a partir de la evidencia, esas reglas son:

- a) Regla de realidad, una posibilidad que es real no puede ser apropiadamente ignorada.
- b) Regla de creencia, una posibilidad que se cree puede ser obtenida no puede ser apropiadamente ignorada.
- c) Regla de la confiabilidad, una posibilidad que no sea obtenida por medios fiables debe ser apropiadamente ignorada.
- d) Regla de semejanza, una posibilidad que tiene parecido a otra que no puede ser apropiadamente ignorada, no puede ser apropiadamente ignorada.
- e) Regla de atención, una posibilidad que no es ignorada, no puede ser apropiadamente ignorada.

De esta manera, una forma de cerrar la brecha de contextualismo epistémico entre investigadores que usan las técnicas de *big data* y la ciencia de datos, y de quienes aún son escépticos ante su lenguaje, método y técnicas, sería mediante el aseguramiento de que algunas posibilidades, más allá de lo que la comunidad científica acepta, sean rechazadas o estén en duda frente a una investigación que utiliza el *big data* y sus técnicas, y esto tendría relación con lo que para unos y otros deben ser las características de la evidencia en el ámbito de las investigaciones con *big data*. De esta manera, debiera considerarse, en línea con las reglas de Lewis, que:

- i) Regla de realidad: Los datos estén siempre definidos en términos relacionales y representativos. Representativos en cuanto a que los datos deben ser una representación confiable de la realidad, una fuente de información empírica obtenida de la interacción con el mundo; y relacionales en cuanto a que deben ser interpretados a partir de una red de relaciones conceptuales, materiales y sociales que deben ser expresadas de forma explícita para justificar los resultados de algún análisis. En



consecuencia, es indispensable que un investigador esté atento, tenga claras, y pueda demostrar las circunstancias en las que se han producido los datos o la forma en que han sido trasladados a lo largo del tiempo, antes de considerar su utilización (Leonelli, 2019).

- ii) Reglas de confiabilidad y creencia: El mantenimiento regular de las infraestructuras que soportan las tecnologías de la información es necesario para justificar la confianza sobre *big data*: La acumulación y la interoperabilidad de los datos requiere de un enorme aparato conceptual, material e institucional bajo la forma de infraestructuras, de bancos de datos, reglamentaciones y programas de formación adaptados. Estos aparatos requieren ellos mismos, financiamientos específicos dedicados a mantener y a tener siempre actualizados sus elementos a lo largo del tiempo (Leonelli, 2019). De esta manera se crea un sistema de creencias y confianza común, en la comunidad científica y entre esta y los usuarios de *big data* y ciencia de datos.
- iii) Regla de semejanza: La infraestructura y las competencias en gestión de datos son esenciales para la extracción de conocimiento a partir de los metadatos: Es crucial que los investigadores implicados en el análisis de *big data* se interesen por el funcionamiento de los bancos de datos y de los algoritmos utilizados para movilizar y analizar los datos, con el fin de poder evaluar de forma crítica el impacto de esos instrumentos, las metodologías que están relacionadas, y los sistemas de clasificación sobre la información extraída a partir de los datos. Así mismo, los investigadores de todos los dominios deben recibir una educación mínima para la utilización de tecnologías, infraestructura y métodos relativos al análisis de la *big data* y la ciencia de datos.
- iv) Regla de atención: La elección de los datos a insertar en un banco de datos debe estar documentada de forma

explícita y respaldada por las buenas prácticas utilizadas desde siempre por las ciencias naturales y sociales para la recolección, tabulación, almacenamiento y procesamiento de los datos, de tal forma que se pueda demarcar claramente las posibilidades en las cuáles esos datos pueden ser generadores de evidencia fiable y en las cuáles no pueden serlo.

## Conclusiones

El progreso tecnológico de las computadoras y las tecnologías de la información ha cambiado el mundo en el que vivimos creando un contexto diferente al que existía en el pasado. Este cambio de contexto ha tenido impacto en diversos ámbitos sociales. La *big data* y los algoritmos de la ciencia de datos facilitan la producción de conocimiento, a partir de las nuevas técnicas de recolección, almacenamiento y procesamiento de datos. Así mismo, la popularización de estas técnicas ha abierto espacio a nuevos participantes, fuera de la comunidad científica, que postulan descubrimientos a los que atribuyen como científicos a partir de una secuencia de acciones determinadas que les permiten reconocer patrones, y que no necesariamente se enmarcan dentro del método científico tradicional generando que hayan disputas entre quienes creen que la *big data* y las técnicas de la ciencia de datos han llegado para convertirse en la herramienta más relevante para crear conocimiento de la realidad, y quienes creen que no. En ese ámbito, es posible describir la existencia de dos contextos epistémicos enfrentados que sin embargo pueden reconciliarse en cuanto a saber si se satisfacen las reglas de exclusión de posibilidades planteadas por Lewis en su artículo *Ellusive Knowledge*.

## Referencias bibliográficas

- Ahonen, P. (2015). Institutionalizing Big Data methods in social and political research. *Big Data and Society*, 2(2), 1–12. <https://doi.org/10.1177/2053951715591224>
- Baumann, P. (2016). *Epistemic Contextualism A Defense*. Oxford University Press.
- Conway, D. (2010). The Data Science Venn Diagram. Retrieved from <http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- Davenport, T. H., & Prusak, L. (1997). *Information ecology*. Oxford University Press.
- Driscoll, K., & Walker, S. (2014). Working within a black box: Transparency in the collection and production of big twitter data. *International Journal of Communication*, 8(1), 1745-1764.
- Drucker, P. (1988). The Coming of the New Organization. Retrieved from <https://hbr.org/1988/01/the-coming-of-the-new-organization>
- Favaretto, M., de Clercq, E., Schneble, C. O., & Elger, B. S. (2020). What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. *PLoS ONE*, 15(2), 1-20. <https://doi.org/10.1371/journal.pone.0228987>
- Frické, M. (2015). Big Data and Its Epistemology. *Journal of the Association for Information Science and Technology*, 66(April), 651–661. <https://doi.org/10.1002/asi>
- Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *Journal of Strategic Information Systems*, 26(3), 191-209. <https://doi.org/10.1016/j.jsis.2017.07.003>
- Hekler, E. B., Klasnja, P., Chevance, G., Golaszewski, N. M., Lewis, D., & Sim, I. (2019). Why we need a small data paradigm. *BMC Medicine*, 17(1), 1-9. <https://doi.org/10.1186/s12916-019-1366-x>
- Hirschheim, R. (2021). The attack on understanding: How big data and theory have led us astray: A comment on Gary Smith's

- Data Mining Fool's Gold. *Journal of Information Technology*, 36(2), 176-183. <https://doi.org/10.1177/0268396220967677>
- Ichikawa, J. J. (2017). *The Routledge Handbook of Epistemic Contextualism. The Routledge Handbook of Epistemic Contextualism*. <https://doi.org/10.4324/9781003043645>
- Jones, M. (2019). What we talk about when we talk about (big) data. *Journal of Strategic Information Systems*, 28(1), 3-16. <https://doi.org/10.1016/j.jsis.2018.10.005>
- Lee, A. J., & Cook, P. S. (2020). The myth of the “data-driven” society: Exploring the interactions of data interfaces, circulations, and abstractions. *Sociology Compass*, 14(1), 1-14. <https://doi.org/10.1111/soc4.12749>
- Leonelli, S. (2019). Cinq façons dont les données massive nuisent à la.
- Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy*, 74(4), 549-567. <https://doi.org/10.1080/00048409612347521>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work and Think*.
- Mazzocchi, F. (2015). Could Big Data be the end of theory in science? *EMBO Reports*, 16(10), 1250-1255. <https://doi.org/10.15252/embr.201541001>
- Obar, J. A. (2020). Sunlight alone is not a disinfectant: Consent and the futility of opening Big Data black boxes (without assistance). *Big Data and Society*, 7(1). <https://doi.org/10.1177/2053951720935615>
- Rysiew, P. (2021). Epistemic Contextualism.
- Sanjuán, M. (2021). On upholding empirical evidence. McIntyre in the Scientific Attitude. *Daimon*, 0507(82), 189-195. <https://doi.org/10.6018/DAIMON.436061>
- Slout, B., Broeders, D., & Schrijvers, E. (2016). Exploring the Boundaries of Big Data. *Exploring the Boundaries of Big Data*. <https://doi.org/10.5117/9789462983588>
- Succi, S., & Coveney, P. V. (2019). Big data: The end of the scientific method? *Philosophical Transactions of the Royal Society A*:

*Mathematical, Physical and Engineering Sciences*, 377(2142).  
<https://doi.org/10.1098/rsta.2018.0145>

van Benthem, J., & Pacuit, E. (2011). Dynamic Logics of Evidence-Based Beliefs. *Studia Logica*, 99(1), 61–92. <https://doi.org/10.1007/s11225-011-9347-x>

Van Der Sloot, B., Broeders, D., & Schrijvers, E. (2016). *Exploring the Boundaries of Big Data*.

