

La utilidad del Big Data en las estadísticas públicas y empresas privadas

The utility of Big Data in public statistics and private companies

Fernando Ariel Manzano [I]
Daniela Avalos [II]

Resumen

El crecimiento exponencial del almacenaje y explotación de los datos, denominado Big Data, dado por las tecnologías de mediación digital, enfrenta fuertes limitaciones. El problema de la calidad de los datos es un desafío para las organizaciones, privadas y públicas. En las primeras, se resuelven en función del valor generado para la actividad económica de las empresas. En las últimas, se vincula al trabajo de los institutos de estadística, abocados a prevenir las fuentes de invalidez. Este estudio de carácter exploratorio, compara el impacto de los macrodatos en ambos entornos. Se propone hallar las diferencias en las distintas dimensiones de la gestión de los datos entre los ámbitos públicos y privados, tomando como factor relevante la calidad de los datos.

Palabras claves: privacidad; estadísticas oficiales; estadísticas privadas; Big Data; calidad de los datos

Abstract

The exponential growth of data storage and exploitation, called Big Data, enabled by digital mediation technologies, faces strong limitations. The problem of data quality is a challenge for private and public organizations. In the former, it is resolved based on the value generated by the economic activity of the companies. In the latter, it is linked to the work of statistical institutes, aimed at preventing sources of disability. This exploratory study compares the impact of Big Data on both environments. The intention is to find differences in distinct dimensions of data management between public and private spheres, taking data quality as a relevant factor.

Keywords: *privacy; official statistics; private statistics; Big Data; data quality.*



Introducción

En el contexto de la denominada Revolución 4.0 y de la irrupción de las tecnologías inteligentes – Smart technologies – como el internet de las cosas o el blockchain, la Inteligencia Artificial,¹ entre otras tecnologías de mediación digital (Serna, 2021). Han dado lugar en la última década a un incremento exponencial del almacenaje y explotación de los datos, también llamados Big Data,² que los usuarios consumen pero que también producen (Salvador y Ramió, 2020). Se ha convertido en un reto para las organizaciones, tanto públicas como privadas, para aprovechar en forma acertada y eficiente esta explosión de datos (Ontiveros y López Sabater, 2017).

Los avances de Big Data y de sus disciplinas asociadas, provienen fundamentalmente desde el sector privado, motivados por producir algoritmos que generen información con valor económico. Estas ventajas técnicas, no están exentas de limitaciones conceptuales y operativas, como la ausencia de disponibilidad de información contrafáctica (Pérez-Rave et al., 2019; Sosa Escudero, 2020). No obstante, las empresas están dispuestas a sacrificar exactitud o precisión a cambio de obtener un gran nivel de detalle de los macrodatos con el objetivo de encontrar información que permita maximizar la generación de valor económico (Garzón Arredondo, 2015).

Los gobiernos están haciendo un esfuerzo para gestionar estos grandes datos, como el Big Data en beneficio público (Galimany Suriol, 2015). El principal desafío es asegurar que la información generada sea de calidad, imparcial y su difusión alcance a todos

los agentes en forma simultánea (Ábrego Molina y Flores Mérida, 2021). Debe tenerse en consideración, que la estadística nacional representa un bien público estratégico. Esta valoración se sustenta en el trabajo de los institutos de estadística, abocados a prevenir las diferentes fuentes de invalidez que afectan la calidad de los datos³ y orientar el modo de corregirlos⁴ (Naciones Unidas, 2017; Giusti y Massé, 1997). Siendo de gran importancia realizar comprobaciones de coherencia en conjunto con los especialistas temáticos (Naciones Unidas, 2011).

Este estudio se propone comparar el impacto del Big Data en las empresas privadas y los institutos de estadísticas nacionales, a partir de una amplia búsqueda de información especializada. El análisis desarrollado, de carácter exploratorio, describe las diferencias en los diseños metodológicos, objetivos o resultados, riesgos, regulaciones, participantes, entre otros elementos que cobran diferentes valoraciones en el entorno de la producción de las estadísticas públicas en comparación al sector empresarial privado. Como consecuencia, estos elementos ofrecen la posibilidad de explorar la valoración de la calidad de los datos que operan en los entornos públicos y privados.

La calidad de los datos en las estadísticas oficiales

Si bien no existe una definición de calidad de la información estadística acordada internacionalmente⁵ tradicionalmente se suelen mencionar tres aspectos principales: pertinencia (la validez de la

forma de medición), exactitud (la fiabilidad o confiabilidad de la estimación) y disponibilidad (la claridad y confiabilidad de la información para los usuarios) (Pettersson, 1992). Sin embargo, en las instituciones nacionales de estadística la calidad hace referencia principalmente a la exactitud⁶ (Viggo et al., 2003). En tal sentido, hace algunas décadas las oficinas de Estadística, no brindaban determinada información a los usuarios, debido a diferentes dificultades, como ser: no poseer la calidad adecuada, la muestra necesaria, la debida consistencia, entre otras.

En 1994, la Organización de las Naciones Unidas (ONU) emitió los Principios Fundamentales de las Estadísticas Oficiales. En los cuales se expresan las distintas dimensiones, cada dimensión posee características y requerimientos propios,⁷ que componen la calidad estadística. A partir de estos principios, el concepto tradicional de calidad fue cambiando,⁸ haciéndose gradualmente más comprensivo,⁹ se modificó su amplitud, enfocándose en siete criterios: relevancia, exactitud, oportunidad y puntualidad, accesibilidad y claridad, comparabilidad, coherencia y completitud, y la satisfacción de las expectativas de los diferentes tipos de usuarios (Elvers, 2002; Naciones Unidas, 2004).

Las nuevas fuentes de información, como el Big Data deben ser adecuadamente adaptadas para ser de utilidad en la elaboración de las estadísticas oficiales. Así como también estas últimas, se han puesto nuevos retos desde hace más de una década, enmarcados en el proceso de modernización e industrialización de la producción estadística oficial (Salgado, 2017).

En el año 2014 se realizó en Beijing la primera Conferencia Internacional sobre Big Data en las Estadísticas Oficiales.¹⁰ En la cual se creó un Grupo de Trabajo Mundial de Naciones Unidas, para llevar adelante un programa mundial de utilización de Big Data en Estadísticas Oficiales, se consensuó que su labor se dirigiría principalmente hacia el marco de calidad de los macrodatos¹¹ (Naciones Unidas, 2015). El concepto ampliado de calidad en los sistemas estadísticos nacionales¹² incluye una gama amplia de formatos y fuentes de datos no recopilados con fines estadísticos, como son los datos administrativos o las fuentes Big Data (Linerós, 2017), este se entiende cada vez más como un concepto multidimensional¹³ (Naciones Unidas, 2010).

Teniendo en cuenta este criterio, los productores de estadísticas oficiales seleccionan las fuentes de obtención de datos, censos, encuestas estadísticas,¹⁴ registros administrativos,¹⁵ Big Data, entre otros.¹⁶ Se debe contemplar en la recogida de la información, el grado de calidad, la oportunidad, el nivel de precisión, el costo y la carga que impondrán¹⁷ (Lindenboim, 2010). También, es responsabilidad de los institutos de estadística informar a los usuarios a través de los metadatos, el grado de integridad, precisión, validez y confiabilidad de los datos generados según, por un lado, cual sea el nivel y las características del error cometido y por otro, qué criterios se siguieron para tratar de corregirlos¹⁸ (Giusti y Massé, 1997; Cepal, 2020).

En la etapa preparatoria de un censo o una encuesta pueden acarrear conceptos no claros o erróneos. Luego de la realización

de un operativo estadístico, en el proceso de evaluación se detectan diferentes errores, como ser: omisiones o duplicaciones de unidades de relevamiento, sean estas de población, vivienda, hogar; errores de los encuestadores; no respuestas de preguntas; pérdidas de formularios; registros tardíos; mala declaración de la edad; inconsistencias entre variables de un mismo registro o entre variables de diferentes registros pertenecientes a una misma unidad de relevamiento; discrepancias respecto de información que proveen fuentes de datos externas; errores de los declarantes; equivocaciones en la codificación y la introducción de datos; errores en la revisión manual o informatizada de los datos; tabulaciones erróneas de los resultados; entre otros. El proceso de revisión debe tratar de conservar los originales en la mayor medida posible (Cepal, 2011; Naciones Unidas, 2011).

El ciclo de vida del dato en la estructura Big Data

En la actualidad, el uso masivo de dispositivos móviles, el internet de las cosas, y la economía digital, conlleva a que prácticamente toda actividad genere datos digitales. Teniendo en cuenta que los dispositivos se conectan entre sí e intercambian información continuamente y adicionando el incremento del número de usuarios con el paso del tiempo, resulta inevitable un aumento del flujo de datos que se manipulan diariamente, representando una oportunidad para las empresas (Vásquez Valdivia, 2021). En relación a este objetivo, las empresas privadas respondieron de diversas maneras, implementando un gobierno de

datos (Garzón Arredondo, 2015), usando herramientas de analítica de datos, entre otros procesos destinados a extraer valor a la gran cantidad de datos (Klatt et al., 2011)

En este sentido, es posible hacer un seguimiento personalizado de las distintas prácticas sociales, por ejemplo, las transacciones comerciales (Martínez, 2019), de los usuarios en internet, y luego codificarlas y sistematizarlas en una base de datos. En este contexto, los grandes volúmenes de datos son capturados y clasificados para después ser comercializados y controlados en un proceso denominado datificación de la realidad (Molina y Mérida, 2021).

Cabe destacar que no todos los datos son accesibles debido a los altos costos o a las propias políticas de encriptación informática, las empresas se protegen codificando la información de manera que no pueda ser interceptada (Chávez, 2019). Solo las compañías que prestan los servicios de Internet tienen acceso a la totalidad de los datos (Rojo y Sánchez, 2019; Manovich, 2012). El objetivo final de una empresa privada es maximizar la creación de valor mientras se minimizan los costos, generando ventajas competitivas. En el marco del Big Data, se agrupan un conjunto de procesos para lograr monetizar la gran cantidad de datos (Cruz Bolívar, 2017). Existen múltiples versiones de la cadena de valor del dato, entre las más habituales cabe mencionar:

1) Fase de provisión del dato: se releva la información que genera la empresa, y luego se combina con información externa, de fuentes muy heterogéneas (Galimany Suriol, 2015). La forma de capturar la información depende fundamentalmente del tipo de datos, estructurados,¹⁹ semiestructurados²⁰

y desestructurados,²¹ la frecuencia en que se releva, en tiempo real, *near real time* o *batch*, y la exactitud y calidad de los mismos (Molina y Mérida, 2021). Los arquitectos de datos son los encargados de seleccionar la tecnología adecuada y los ingenieros de datos los responsables de construir el *software* para integrar la información de los sistemas fuente en la plataforma Big Data, de manera que pueda ser utilizada de forma correcta por los científicos de datos.²² Cabe destacar que la confiabilidad de los datos puede ser violada en cualquier punto a lo largo de su ciclo de vida (Garzón Arredondo, 2015).

2) Fase de transformación: Una vez almacenada la información en el sistema de Big Data, es necesario proceder a limpiarla y enriquecerla con el objetivo de proveer información precisa y de calidad, de confianza, a los usuarios (Galimany Suriol, 2015). Siendo los repositorios de datos son tan desproporcionadamente grandes que resulta imposible tratarlos con las herramientas de bases de datos y analíticas convencionales (Esumer, 2017 não consta). Los científicos de datos dedican un 80% de su tiempo al proceso de limpieza y almacenamiento de datos, proceso conocido como *data wrangling* o *data munging*, quedando en segundo lugar lo que debería ser su actividad principal: la analítica avanzada (Mons, 2020). El proceso de *data wrangling*,²³ se realiza durante el pre-procesamiento de datos, comprendiendo las siguientes tareas: edición, limpieza y transformación de los datos, ajuste para detectar omisiones, verificación

de legibilidad, anomalías y consistencia para la codificación y almacenamiento (Ramírez-Morales et al., 2018).

3) Fase de descubrimiento y modelado: proceso de análisis por parte de los analistas de negocio y los científicos de datos con el objetivo de extraer el máximo valor transformando los datos en información. La analítica de datos consiste en extraer conocimiento de la información para responder a las preguntas que se le plantean en los diferentes casos de uso. El análisis de información en redes sociales es de gran potencial para conocer el comportamiento del consumidor, pero al tratarse de datos no estructurados, su principal inconveniente viene dado por el alto costo de su procesamiento (Cruz Bolívar, 2017).

4) Fase de exposición: Dado que no siempre es posible extraer valor de la información generada. Para minimizar esta pérdida, las tres fases anteriores se realizan de la manera más económica. El principal inconveniente se presenta en la manipulación de estos datos en los entornos analíticos. Debido a ello ha surgido *chief data officer* (CDO), que es el ejecutivo responsable de gobernar todos los datos de la empresa, y de definir las diferentes políticas de gobierno en relación con la estrategia, control y explotación eficiente de la información con el objetivo de crear valor para el negocio. Recién en esta fase se realizan las tareas de validación de premisas asumidas anteriormente respecto a los datos de entrada, documentación del glosario de negocio con los nuevos términos, entre otros, se industrializa todo el proceso anterior.

Complejidades y desafíos en la gestión pública para el manejo de macrodatos

En los últimos años, el uso intensivo de las tecnologías de la información y de las comunicaciones (TIC), aumentó la demanda de los usuarios de la apertura de datos para ser reutilizados, conduciendo al surgimiento de los Datos Públicos Abiertos²⁴ (DA) u open data, por su denominación en inglés (Christodoulou et al., 2018 não consta).

Muchas nuevas fuentes de datos digitales, no se adecuan a las responsabilidades de las Estadísticas Oficiales de los países. Estas últimas, en el marco de la gestión de DA, deben velar por la calidad de los datos y su uso adecuado, cumpliendo con los estándares de calidad,²⁵ confidencialidad²⁶ y seguridad (Vásquez Valdivia, 2021; Salvador y Ramió, 2020), cobrando especial importancia atender a la evidencia, la seguridad y privacidad del contenido de los nuevos datos (Christodoulou et al., 2018 não consta). Además, dado los principios éticos elementales del sector público, debe garantizar transparencia e identificar la existencia de errores²⁷ de estos grandes volúmenes de datos (Paliotta, 2018).

Un instrumento utilizado en los procesos de análisis de datos es el aprendizaje automático, este procedimiento hace posible que los sistemas identifiquen o aprendan patrones de manera autónoma a partir de un conjunto de datos. Así como también realizar predicciones sobre aspectos de la sociedad y tomar decisiones sin la necesidad de programar estas tareas de forma explícita

(Escobar Gutiérrez et al., 2021). A partir de esto surge una preocupación reiterada, respecto a la calidad de los datos, considerando que la utilización de datos de baja calidad²⁸ dificulta la producción de resultados fiables (Wirtz et al., 2019; Susar y Aquaro, 2019; Gerbert et al., 2017). Además, muchos académicos de las ciencias sociales se posicionan críticamente en relación a la posibilidad de hallar soluciones a problemáticas sociales mediante procesos tecnológicos, sin necesidad de intervención de las teorías sociales (Boellstorff, 2013; Bollier, 2010; Markham, 2013; Burrell, 2012). Así como también rechazan la capacidad de estimar el futuro a partir de grandes volúmenes de datos, de forma determinista, como simple prolongación lineal del pasado (Del-Fresno-García, 2014). Para estos autores el Big Data no tiene pretensión de un conocimiento con fuerza normativa, simplemente se restringe a tomar decisiones acotadas y situadas (Becerra, 2018).

Un gran número de investigaciones remarcan que las fuentes de información de una estructura Big Data, contienen diversas dificultades, muchas de ellas vinculadas entre sí, como la falta de evidencia del fenómeno relevado (Luo et al., 2019), la baja calidad de las fuentes de información,²⁹ poseen una elevada proporción de datos no estructurados y erróneos³⁰ (Paliotta, 2018), que se expresa en la poca usabilidad de las mismas (Pérez-Rave et al., 2019), la metodología estadística para realizar inferencias respecto de las poblaciones de interés (Salgado, 2017), la preeminencia del volumen de información por sobre la precisión analítica (Del-Fresno-García, 2014), las restricciones en el acceso institucional a los datos, cambios legales en los países. Además de los problemas de

duplicación, errores tipográficos, información falsa o basada en percepciones subjetivas de una arquitectura de Big Data (Monleon-Getino, 2015; Rodríguez et al., 2017). En este sentido, cobra relevancia la generación de estudios de viabilidad convincentes para que la comunidad estadística se posicione de forma unificada en relación al uso de macrodatos.

Discusión

En el caso de las estadísticas oficiales, se debe lidiar con la baja calidad de las fuentes de información de una arquitectura de Big Data (Pérez-Rave et al., 2019), debido al bajo nivel de estructuración de los datos,³¹ que permitan asociarlos a una población o fenómeno de referencia, la diversidad de contenidos y fuentes (Monleon-Getino, 2015), la espontaneidad y correspondencia con poblaciones heterogéneas, lo contrario de la independencia y distribución idéntica de los tradicionales.

La analítica Big Data³² viene siendo considerada como una forma de análisis cuantitativo más desarrollada, que destaca como novedades: recopilación de datos indiscriminada y técnicas de agregación y cruce de datos (Baruh y Popescu, 2015), permitiendo inferir información que los datos no contienen de forma explícita (Tufekci, 2015). Los Datos de Analítica Web surgen de analizar el tráfico de usuarios en una web de una organización, son tanto cuantitativos como cualitativos. Esta analítica se propone brindar indicadores relevantes que permitan cuantificar aspectos o estados de la página web en función de determinado objetivo estratégico de la empresa (Galimany Suriol, 2015).

A partir de la diferenciación entre cantidad y profundidad de los datos, resulta de interés discriminar entre datos de superficie, *surface data*, y datos profundos, *deep data*. En el primero, se vinculan a metodologías cuantitativas, se aspira a la búsqueda de patrones de regularidad, así como relaciones causales entre variables (Felt, 2016). Los últimos, son utilizados en metodologías cualitativas,³³ mayormente en los trabajos etnográficos, posibilitan conocer con mayor nivel de especificidad los fenómenos sociales, teniendo en cuenta la perspectiva de los sujetos involucrados (Manovich, 2012; Forni y Grande, 2020). Una nueva forma de conocer la actividad social y cultural denominada analítica cultural (*cultural analytics*) o computación social (*social computing*) (Ardèvol, 2016). Estos enfoques actualizan el viejo debate de los métodos cualitativos y los cuantitativos, mediante la hibridación entre análisis automatizados y métodos artesanales (Gindin y Busso, 2018). Se pone énfasis en la significación y en las correlaciones entre variables para llegar a una mayor comprensión del fenómeno estudiado, dejando de lado la representatividad de la muestra³⁴ y la búsqueda de relaciones causales (Zwitter, 2014). Destacando la posibilidad de documentar casi la totalidad de datos sobre un acontecimiento insospechado, para hallar eventos verdaderamente relevantes, que no serían percibidos desde una perspectiva que se limita al uso de muestras representativas, esto incrementaría el error de validar correlaciones espurias (Garzón Arredondo, 2015). Se debe tener presente, que estos “datos culturales” relevados mediante el análisis Big Data, no se traducen en patrones culturales, tan solo es posible darle sentido

desde un análisis cultural. A modo de ejemplo, a ningún analista de Big Data se le ocurriría buscar patrones significativos en millares de distintas poblaciones de bacterias sin tener al lado un biólogo que guiara sus codificaciones e interpretara sus hallazgos (Ardèvol, 2016, p. 32).

Conclusiones

El incremento exponencial del almacenaje y explotación de los datos de los últimos años se ha convertido en un reto para las organizaciones públicas y privadas para aprovechar de acuerdo a los objetivos y requisitos que cada una de ellas promueven.

Las empresas privadas son las principales productoras y recolectoras de la industria Big Data, usan los grandes volúmenes de datos para mejorar la toma de decisiones, generar nuevos productos o servicios, entre otras potencialidades económicas. Estos datos capturados para ser comercializados y controlados, están expuestos en cualquier fase de procesamiento a perder confiabilidad. No obstante, las organizaciones privadas están dispuestas a sacrificar exactitud o precisión mientras se genere beneficio económico.

El rol de los gobiernos, como poseedor y publicador de datos públicos con mayor apertura cubriendo la necesidad de los usuarios, fue ampliando el concepto

tradicional de la calidad de la estadística. A nivel internacional la utilización de Big Data en Estadísticas Oficiales, se concentra principalmente en el marco de calidad de los macrodatos, siendo de gran importancia realizar comprobaciones de coherencia junto a especialistas temáticos.

En las estadísticas del sector privado se dedica la mayor parte del tiempo a limpiar y enriquecer la información. Esta tarea se realiza de la manera más económica, siendo el objetivo principal maximizar las ganancias empresariales. Por tanto, el principal inconveniente viene dado por el alto costo de este procesamiento.

En el caso de las estadísticas oficiales, el valor depende de su capacidad de crear conocimiento en la sociedad. Su difusión debe realizarse en forma simultánea a todos los usuarios, satisfaciendo la demanda de los mismos por mayor apertura de datos, cobrando especial importancia atender a la privacidad. Además, se deben informar a los usuarios el grado de validez y confiabilidad de los datos disponibles, tanto de los errores cometidos, como los criterios adoptados para tratar de corregirlos.

Un punto en común en las organizaciones privadas y públicas, es el costo y la carga que imponen los distintos tipos de fuentes, siendo este un factor de relevancia en la elección de las mismas. Asimismo, no todos los datos son accesibles debido a los altos costos o a las políticas de encriptación.

[I] <https://orcid.org/0000-0002-1513-4891>

Universidad Nacional del Centro de la Provincia de Buenos Aires. Consejo Nacional de Investigaciones Científicas y Técnicas. Buenos Aires, Argentina.

fernandoarielmanzano@fch.unicen.edu.ar

[II] <https://orcid.org/0000-0002-0762-8437>

Universidad de Tres de Febrero, Programa de Posgrado en Estadística. Buenos Aires, Argentina.

danielasoledad.av@gmail.com

Notas

- (1) La IA aún no posee una definición universalmente aceptada (Grosz et al., 2016). En una aproximación amplia, el concepto se asocia a sistemas de computación que recaban información de diferentes fuentes, con capacidad para automatizar actividades, realizar operaciones análogas al aprendizaje y toma de decisiones, y evolucionar con o sin la intervención humana (Castro y New, 2016).
- (2) Este término se ha acuñado en español como Macrodatos en algunas traducciones (Naciones Unidas, 2015).
- (3) La producción de información está sujeta a la existencia de errores de diversa índole y magnitud. Los mismos pueden introducirse en diferentes momentos del proceso de producción (Campos y Zambrano, 2020). Los errores según el tipo de invalidez pueden ser aleatorios o sistemáticos. Los errores aleatorios se producen sin ningún orden lógico, afectando a todos o casi todos los registros de manera uniforme, no distorsionando la distribución de los valores. Los sistemáticos provienen mayormente de un mal entendimiento de las preguntas, conceptos o instrucciones, de cualquiera de los individuos que participan en las distintas etapas del relevamiento, suelen afectar a un grupo específico de variables o registros, y generan cierta distorsión en la distribución de los valores (Indec, 1987).
- (4) Según los resultados de la última encuesta sobre el estado de avance de los censos de 2020, la identificación de las necesidades nacionales, el control de calidad y cobertura es uno de los temas prioritarios donde los países requerían asistencia técnica. Cabe destacar que el 75% respondió que realizaría cambios en sus controles, debido a que este factor se relaciona a la credibilidad y la confianza de la organización (Cepal-Celade, 2020).
- (5) Una de las instituciones pioneras en el aseguramiento de la calidad de la estadística pública fue Statistics Sweden, quienes en 1992 establecieron una de las primeras definiciones de calidad en la información estadística.
- (6) Como tener en cuenta los errores, de muestreo o de otro tipo, que influyen en el valor de las estimaciones, y los intervalos basados en los conocimientos que podrían servir de base para las declaraciones de confianza (Naciones Unidas, 2010).
- (7) Las acciones que se toman para tratar alguna de las dimensiones de calidad pueden afectar a otra dimensión, en formas completamente impredecibles. Para lograr un balance apropiado de las dimensiones de la calidad y otros factores se deben tomar decisiones basadas en el conocimiento, la experiencia, las revisiones, la retroalimentación, la consulta e, inevitablemente, el juicio personal (Naciones Unidas, 2004).

- (8) Prácticamente se suelen brindar respuesta a todos los pedidos, salvo, aquellos casos que pongan en cuestión el secreto estadístico. Se deben informar cuando corresponda, las eventuales limitaciones de la información brindada (Cepal, 2000).
- (9) Aun cuando los datos sean exactos, no tienen la calidad suficiente si se presentan tan tarde que dejan de ser útiles, o no son de fácil acceso o están en contradicción con otros datos creíbles o su obtención resulta demasiado costosa (Naciones Unidas, 2010, p. 59).
- (10) La segunda Conferencia Internacional realizada en 2015 en Abu Dabi, Emiratos Árabes Unidos. En la tercera Conferencia Global, realizada en Dublín, Irlanda en 2016. En la cuarta Conferencia Global, realizada en Bogotá, Colombia en noviembre de 2017.
- (11) Los proyectos de macrodatos están conformados por agentes del sector privado y de institutos de investigación.
- (12) La División de Estadísticas de la ONU tiene en cuenta que la calidad de las estadísticas públicas y, en consecuencia, la calidad de la información que dispone el gobierno, la economía y el público, dependen en gran medida de la colaboración de los ciudadanos, las empresas y otras fuentes que deben proporcionar datos pertinentes en forma veraz y oportuna (Jiménez Mora, 2009).
- (13) Un dato obtenido mediante un procedimiento estadístico puede ser exacto, pero no se puede decir por ello que es de calidad si se genera demasiado tarde para su uso, o si es de difícil acceso, o si es contradictorio con otra información estadística producida (Matus, 2007).
- (14) Encuesta estadística significa la recolección principal de datos individuales de los informantes de una determinada población, realizada por un productor de estadísticas oficiales con fines exclusivamente estadísticos mediante el uso sistemático de la metodología estadística (Cepal, 2020, p. 10).
- (15) Registros administrativos significa datos recogidos por, o en nombre de las autoridades nacionales y locales, incluido el banco central, sean o no un productor de estadísticas oficiales, para fines administrativos, de conformidad con fundamentos jurídicos diferentes de los de la legislación estadística (Cepal, 2020, p. 10). Si bien al principio se consideraba que los datos administrativos no eran adecuados para fines estadísticos, en la actualidad los estadísticos están casi obligados a agotar todas las posibilidades que ofrecen los datos administrativos disponibles antes de realizar una nueva encuesta.
- (16) Siendo necesario considerar todas las fuentes de datos de manera conjunta desde un enfoque multisectorial (Bussi et al., 2017).
- (17) Los bajos resultados en la calidad de los censos de los años ochenta del siglo pasado en la región de América Latina, echaron por tierra el convencimiento de mejora en el tiempo de los operativos, debido al adelanto tecnológico (Arretx y Chackiel, 1985). Posteriormente se han experimentado mejoras en la calidad de la información recolectada en los censos, encuestas y registros de estadísticas vitales, no obstante, aún hay un déficit grande en la región en todas estas cuestiones (Cavenaghi, 2012).
- (18) Cabe destacar que los organismos internacionales y regionales, aceptan la aplicación de criterios estadísticos y/o demográficos para la asignación de datos faltantes, solo en el caso que el porcentaje de datos faltantes no sea elevado. Los criterios teóricos que sustentan las pautas de consistencia y corrección automática adoptada deben quedar registradas, así como también deben realizarse evaluaciones acerca de los efectos generados en la exactitud de los resultados (Naciones Unidas, 1980).

- (19) Los datos estructurados son especificados con detalle y tienen un esquema y estructura fijada con anterioridad, permitiendo su almacenaje en las tradicionales bases de datos relacionales, uso de tablas, registros y columnas (Galimany Suriol, 2015). Representaban al inicio de esta década apenas el 5% de los datos totales (Cukier, 2010). Ganando relevancia recientemente con las legislaciones de transparencia en las administraciones públicas (Kischinhevsky, 2020)
- (20) Los datos semiestructurados, surgen con el auge de las páginas webs, carecen de estructura, pero poseen un flujo lógico que permiten identificar la actividad en Internet de un ordenador (Jareño y Miralles, 2017). Están conformados principalmente por archivos HTML, XML, y otros formatos similares (Joyanes Aguilar, 2016).
- (21) Este tipo de dato no posee estructura, incluye textos, imágenes, audios, vídeos, posts de Twitter, mails, mensajes, entre otros (Galimany Suriol, 2015; Jareño y Miralles, 2017). Existe cierto desconocimiento en la posibilidad de extraer información válida de los datos no estructurados.
- (22) El científico de datos debe poseer, además de conocimientos de ciencia aplicada, aptitudes para trasladar a los responsables de la toma de decisiones los insights obtenidos durante el proceso de análisis. Tiene que contar además con capacidades de storytelling y visualización de datos, esto genera dificultad en las empresas para encontrar perfiles que reúnan estas habilidades.
- (23) Un concepto novedoso es el denominado Garbage in, gospel out, que implica aceptar ciegamente la información de salida informatizada aun cuando provenga de datos con baja calidad. Asumir estos niveles de errores estadísticos es una de las características distintivas del Big Data (Galimany Suriol, 2015).
- (24) La demanda de la ciudadanía por la apertura de datos públicos, conlleva a que los gobiernos respondan con mayor transparencia, legitimidad, acceso a la información y mecanismos efectivos de rendición de cuentas a la sociedad (Ruvalcaba-Gómez, 2020; Concha y Naser, 2012).
- (25) Se espera que los datos públicos sean de buena calidad, asimismo deben encontrarse estandarizados y con metadatos (información asociada) acerca de cómo se han producido (Vásquez Valdivia, 2021).
- (26) La información asociada a datos personales que permitiría identificar a una persona concreta, requiere una protección especial en ámbitos relacionados con la transparencia, el consentimiento, la calidad, la exactitud, entre otros (Ontiveros y López Sabater, 2017). En este sentido, respecto de la calidad y exactitud de la información, por ejemplo, la red social YouTube enfrenta acusaciones de desinformación en relación a las vacunas para el COVID-19 (Hussein et al., 2020; Tandoc Jr et. al., 2017).
- (27) Problemas de duplicación, errores tipográficos, información falsa o basada en percepciones subjetivas (Rodríguez et al., 2017). Asimismo, se considera que los problemas en torno a la selectividad y la representatividad son más comunes cuando se trata de Big Data que cuando se trata de fuentes más tradicionales de datos de los institutos de estadísticas (Dufty et al., 2014).
- (28) La importancia de los datos para el desarrollo de la IA en el sector público, se obstaculiza con la necesidad de revisar la veracidad de los datos, su calidad, su integridad y su trazabilidad, cuestiones vinculadas especialmente a su confiabilidad. El cumplimiento de estos atributos es el punto de partida para la obtención de resultados y soluciones válidas para la mejora de la actuación pública (Serna, 2021).

- (29) Se han añadido nuevas “Vs” a la definición del Big Data: veracidad (la calidad de los datos capturados), variabilidad (manejo de inconsistencias por el cambio en el significado de los datos) y valor (los ingresos o beneficios derivados) (Ontiveros y López Sabater, 2017).
- (30) Se consideró que los problemas en torno a la selectividad y la representatividad son más comunes cuando se trata de Big Data que cuando se trata de fuentes más tradicionales de datos de los institutos de estadísticas (Dufty et al., 2014).
- (31) Acerca del abordaje de la tipología de los datos (ver notas al pie n. 19, 20 y 21).
- (32) Predomina la comprensión de Big Data como un conjunto heterogéneo de técnicas de procesar grandes cantidades de datos digitales, con el objetivo de extraer conocimiento que permita apoyar la toma de decisiones (Ardèvol, 2016).
- (33) Basado en la Teoría fundamentada que Strauss y Corbin (1994, p. 273) definen como: (...) una metodología general para desarrollar una teoría que esté fundamentada en una recogida y análisis sistemático de los datos. La teoría se desarrolla a lo largo de la investigación, mediante una continua interpelación entre el análisis y la recogida de datos.
- (34) Una muestra es la elección de una parte de la población. El muestreo estadístico implica la obtención de una muestra estadísticamente representativa de la población que se inscribe en un proceso de investigación de carácter cuantitativo donde la teoría del muestreo y de probabilidades son elementos importantes definitorios. Existen estrategias de muestreo no probabilístico o cualitativo que demanda la elección de informantes de la investigación (López-Roldán y Fachelli, 2015).

Referencias

- ARDÈVOL, E. (2016). Big data y descripción densa. *Virtualis*, v. 7, n. 14, pp. 14-38.
- ARRETX, C.; CHACKIEL, J. (1985). *Algunos problemas relativos a la recolección de datos demográficos en los censos de población de América Latina en la década de 1980*. Notas de población.
- BARUH, L.; POPESCU, M. (2015). Big data analytics and the limits of privacy self-management. *New Media y Society*, v. 19, n. 4, pp. 579-596. DOI: 10.1177/1461444815614001.
- BECERRA, G. (2018). Interpelaciones entre el Big data y la Teoría de los sistemas sociales. Propuestas para un programa de investigación. *Hipertextos*, v. 6, n. 9, pp. 42-64.
- BOELLSTORFF, T. (2013). Making big data, in theory. *First Monday*, v. 18, n. 10, pp. 1-17.
- BOLLIER, D. (2010). *The promise and peril of Big Data*. Washington, The Aspen Institute.
- BURRELL, J. (2012). The ethnographer's complete guide to Big Data: small data people in a Big Data World (part 1 of 3). Disponible en: <http://ethnographymatters.net/blog/2012/05/28/smalldata-people-in-a-big-data-world/>. Acceso en: 8 nov 2021.

- BUSSI, J.; MARÍ, G. P.; MÉNDEZ, F. (2017). *El desafío del big data en estadísticas oficiales en Argentina*. Rosario, Facultad de Ciencias Económicas y Estadística de la Universidad de Rosario.
- CAMPOS, J. M.; ZAMBRANO, M. Z. (2020). Calidad de los censos tradicionales de población y vivienda: Evaluación de sus etapas. *Matemática*, v. 18, n. 2, pp. 1-26.
- CASTRO, D.; NEW, J. (2016). The promise of artificial intelligence. *Center for Data Innovation*. Disponible en: <https://www.datainnovation.org/2016/10/the-promise-of-artificial-intelligence/>. Acceso en: 4 ago 2020.
- CAVENAGHI, S. (2012). *Estimaciones y proyecciones de población en América Latina: desafíos de una agenda pendiente*. Rio de Janeiro, Asociación Latinoamericana de Población.
- CEPAL – CELADE (2020). Censos de Población y Vivienda, Informe final de resultados de la Encuesta sobre estado de avance de los censos de 2020 e identificación de las necesidades nacionales. Disponible en: https://celade.cepal.org/censosinfo/documentos/INFORME_CENSOS RONDA 2020 LAYCARIBE.pdf. Acceso en: 5 oct 2021.
- CEPAL (2000). *Sexto Taller Regional sobre Indicadores sobre el Desarrollo Social*. Buenos Aires 15 a17 noviembre 2000. Disponible en: <http://hdl.handle.net/11362/20000>. Acceso en: 5 oct 2021.
- _____. (2011). Guía para asegurar la calidad de los datos censales. Serie n. 74. Disponible en: https://repositorio.cepal.org/bitstream/handle/11362/5515/S1100929_es.pdf?sequence=1&isAllowed=y. Acceso en: 4 oct 2021.
- _____. (2020). *Ley Genérica sobre Estadísticas Oficiales para América Latina (LC/CEA.10/8)*. Santiago. Disponible en: https://repositorio.cepal.org/bitstream/handle/11362/45253/1/S2000046_es.pdf. Acceso en: 5 oct 2021.
- CHÁVEZ, C. A. C. (2019). La encriptación de datos empresariales: ventajas y desventajas. *Recimundo*, v. 3, n. 2, pp. 980-997.
- CHRISTODOULOU, P.; DECKER, S.; DOUKA, A. V.; KOMOPOULOU, C.; PERISTERAS, V.; SGAGIA, S.; TSARAPATSANIS, V.; VARDOUNIOTIS, D. (2018). "Data makes the public sector go round". In: PARYCEK, P.; GLASSEY, O.; JANSSEN, M.; SCHOLL, H. J.; TAMBOURIS, E.; KALAMPOKIS, E.; VIRKAR, S. (eds.). *Electronic Government*. EGOV 2018. Lecture Notes in Computer Science (vol. 11.020). Springer. Disponible en: https://doi.org/10.1007/978-3-319-98690-6_19. Acceso en: 5 oct 2021.
- CONCHA, G.; NASER, A. (2012). *Datos abiertos: un nuevo desafío para los gobiernos de la región*. Cepal. Recuperado de <https://www.bivica.org/files/desafio-gobiernos.pdf>
- CRUZ BOLÍVAR, P. (2017). *Uso de Big Data para la toma de decisiones acordes a la estrategia empresarial en el sector retail*. Bogotá, Universidad Libre.
- CUKIER, K. (2010). *The Economist, Data, data everywhere: a special report on managing information*. Disponible en: <http://www.economist.com/node/15557443>. Acceso en: 15 abr 2022.
- DEL-FRESNO-GARCÍA, M. (2014). Haciendo visible lo invisible: visualización de la estructura de las relaciones en red en Twitter por medio del análisis de redes sociales. *El Profesional de la Información*, v. 23, n. 3, pp. 246–252. Disponible en: <https://doi.org/10.3145/epi.2014.may.04>. Acceso en: 8 nov 2021.
- DUFTY, D.; BÉRARD, H.; LEFRANC, S.; SIGNORE, M. (2014). A suggested framework for the quality of big data. *Big Data Quality Framework*, v. 4.01.

- ELVERS, E. (2002). Comparison of Survey and Register Statistics. In: THE INTERNATIONAL CONFERENCE ON IMPROVING SURVEYS. Denmark, University of Copenhagen.
- ESCOBAR GUTIÉRREZ, E.; RAMÍREZ ROA, D. P.; QUEVEDO HERNÁNDEZ, M.; INSUASTI CEBALLOS, H. D.; JIMÉNEZ OSPINA, A.; MONTENEGRO HELFER, P.; ZAPATA, E. (2021). *Aprovechamiento de datos para la toma de decisiones en el sector público*. Bogotá, Corporación Andina de Fomento.
- ESUMER, I. U. (2017). Big Data y los nuevos manejos de la información. *Mercatec*, v. 3, n. 54. Disponible en: <https://repositorio.esumer.edu.co/bitstream/esumer/1905/1/BIG%20DATA.pdf>. Acceso en: 5 oct 2021.
- FELT, M. (2016). Social media and the social sciences: How researchers employ Big Data analytics. *Sage Journal*, v. 3, n. 1, pp. 1-15.
- FORNI, P.; GRANDE, P. D. (2020). Triangulación y métodos mixtos en las ciencias sociales contemporáneas. *Revista mexicana de sociología*, v. 82, n. 1, pp. 159-189.
- GALIMANY SURIOL, A. (2015). *La creación de valor en las empresas a través del Big Data*. Barcelona, Universidad de Barcelona, Grado de Administración y Dirección de Empresas.
- GARZÓN ARREDONDO, A. (2015). *Evolución e impacto de Big Data en empresas grandes de diferentes industrias del sector corporativo en Antioquia*. Doctoral dissertation. Medellín, Universidad Eafit.
- GERBERT, P.; HECKER, M.; STEINHÄUSER, S.; RUWOLT, P. (2017). *Putting artificial intelligence to work*. BCG Henderson Institute. Disponible en: <https://www.bcg.com/de-de/publications/2017/technologydigitalstrategy-putting-artificial-intelligence-work.aspx>. Acceso en: 4 ago 2020.
- GINDIN, I. L.; BUSSO, M. P. (2018). Investigaciones en comunicación en tiempos de big data: sobre metodologías y temporalidades en el abordaje de redes sociales. *Revista adComunica*, 15, pp. 25-43.
- GIUSTI, A.; MASSÉ, G. (1997). Aspectos conceptuales relativos a la evaluación de calidad. *INDEC, Evaluación de la calidad de datos y avances metodológicos Serie J n. 2*. Buenos Aires.
- GROSZ, B. J.; MACKWORTH, A.; ALTMAN, R.; HORVITZ, E.; MITCHELL, T.; MULLIGAN, D.; SHOHAM, Y. (2016). *Artificial intelligence and life in 2030: One hundred years' study on artificial intelligence*. Stanford University. Disponible en: https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fml.pdf. Acceso en: 4 ago 2020.
- HUSSEIN, E.; JUNEJA, P.; MITRA, T. (2020). *Measuring misinformation in video search platforms: an audit study on YouTube*. Proc. ACM Hum.-Comput. Interact., 4(CSCW1). Disponible en: <https://doi.org/10.1145/3392854>. Acceso en: 15 abr 2022.
- INDEC – Instituto Nacional de Estadística y Censos (1987). III Curso de informática para estadísticos. Módulo V. Argentina, Centro regional del IBI para la enseñanza de la informática (CREI)- Centro Interamericano de Enseñanza de Estadística (CIENES), 18 de noviembre al 15 de diciembre de 1987. Ciudad Autónoma de Buenos Aires.
- INSTITUCIÓN UNIVERSITARIA ESUMER (2017). Big Data y los Nuevos Manejos de la Información. *Mercatec*, v. 3, n. 54.
- JAREÑO, J. A. A.; MIRALLES, J. M. P. (2017). El paradigma Big Data y el Aprendizaje Automático. *Actuarios*, n. 40, pp. 10-14.
- JIMÉNEZ MORA, J. M. (2009). *Gestión de la calidad de la estadística pública*. Tesis Doctoral. Veracruz/México, Universidad Veracruzana.

- KISCHINHEVSKY, M. (2020). De las síntesis informativas a los resúmenes para altavoces inteligentes, desafíos al periodismo radiofónico de carácter local. *Estudios sobre el mensaje periodístico*, v. 26, n. 1, pp. 167-175. Disponible en: <https://doi.org/10.5209/esmp.67296>. Acceso en: 15 abr 2022.
- KLATT, T.; SCHLAEFKE, M.; MOELLER, K. (2011). Integrating business analytics into strategic planning for better performance. *Journal of Business Strategy*, v. 32, n. 6, pp. 30-39. Disponible en: <https://doi.org/10.1108/02756661111180113>. Acceso en: 8 nov 2021.
- LINDENBOIM, J. (2010). Las estadísticas oficiales en Argentina ¿Herramientas u obstáculos para las ciencias sociales? *Trabajo y Sociedad*, n. 16, pp. 19-38.
- LINEROS, E. M. (2017). "El trinomio dato-información-conocimiento". In: EDIMEMA. *Manual sobre utilidades del big data para bienes públicos*, pp. 35-48.
- LÓPEZ-ROLDÁN, P.; FACHELLI, S. (2015). *Metodología de la investigación social cuantitativa*. Barcelona, Universitat Autònoma de Barcelona.
- LUO, J.-D.; LIU, J.; YANG, K.; FU, X. (2019). Big data research guided by sociological theory: a triadic dialogue among big data analysis, theory, and predictive models. *The Journal of Chinese Sociology*, v. 6, n. 11. Disponible en: <https://doi.org/10.1186/s40711-019-0102-4>. Acceso en: 22 nov 2021.
- MANOVICH, L. (2012). "Trending: the promises and the challenges of Big Social Data". In: GOLD, M. *Debates in the digital humanities*. Minnesota, University of Minnesota Press.
- MARKHAM, A. (2013). Undermining 'data': a critical examination of a core term in scientific inquiry. *Firstmonday*, v. 18, n. 10. DOI:10.5210/fm.v18i10.4868.
- MARTÍNEZ, N. (2019). *Comunicación e información en la era digital*. Disponible en: https://www.cefc.org.ar/assets/files/barda8_martinez-n.pdf. Acceso en: 16 abr 2022.
- MATUS, C. (2007). *Dimensiones de la calidad según OECD y Eurostat*. Chile/Santiago, Instituto Nacional de Estadísticas.
- MOLINA, V. H. A.; MÉRIDA, A. F. (2021). Datificación crítica: práctica y producción de conocimiento a contracorriente de la gubernamentalidad algorítmica. Dos ejemplos en el caso mexicano. *Administración Pública y Sociedad*, n. 11, pp. 211-231.
- MONLEON-GETINO, A. (2015). El impacto del Big Data en la sociedad de la información. Significado y utilidad. *Historia y Comunicación Social*, v. 20, n. 2, pp. 427-445. Disponible en: https://doi.org/10.5209/rev_HICS.2015.v20.n2.51392. Acceso en: 22 nov 2021.
- MONS, B. (2020). Invest 5% of research funds in ensuring data are reusable. *Nature*, v. 578, p. 491.
- NACIONES UNIDAS (1980). Principios y recomendaciones para la elaboración de Censos de Población y habitación. Serie M n. 67. Nueva York.
- _____ (2004). *Manual de Organización Estadística*. Nueva York, EEUU.
- _____ (2010). *Principios y recomendaciones para los censos de población y habitación*. Revisión 2. Informes Estadísticos Serie M n. 67/Rev.2. Nueva York. Departamento de Asuntos Económicos y Sociales. División de Estadística. Publicación de las Naciones Unidas, n. de venta S.07.XVII.8. Disponible en: https://unstats.un.org/unsd/publication/seriesm/seriesm_67rev2s.pdf. Acceso en: 10 oct 2021.
- _____ (2011). *Manual de revisión de datos de los censos de población y vivienda*. Revisión 1. Departamento de Asuntos Económicos y Sociales. División de Estadística.

- NACIONES UNIDAS (2015). Informe del Grupo de Trabajo Mundial sobre los Macrodatos en las Estadísticas Oficiales. E/CN.3/2016/1.
- _____ (2017). *Desarrollo de la capacidad estadística*. Informe del Secretario General. E/CN.3/2017/30.
- ONTIVEROS, E.; LÓPEZ SABATER, V. (2017). *Economía de los Datos. Riqueza 4.0*. Barcelona, Ariel y Fundación Telefónica.
- PALIOтта, A. P. (2018). Nuevas profesiones y técnicas de web data mining en Argentina: el caso del Data Scientist. *Revista del Centro de Estudios de Sociología del Trabajo*, n. 10, pp. 63-94.
- PÉREZ-RAVE, J.; CORREA MORALES, J. C.; GONZÁLEZ ECHAVARRÍA, F. (2019). Metodología para explorar datos abiertos de accidentalidad vial usando Ciencia de Datos: caso Medellín. *Ingeniare. Revista chilena de ingeniería*, v. 27, n. 3, pp. 495-509.
- PETTERSSON, H. (1992). *Control de calidad en estadísticas procedentes de registros y archivos administrativos*. Euskadi, Instituto Vasco de Estadística.
- RAMÍREZ-MORALES, I.; MAZON-OLIVO, B.; PAN, A. (2018). Ciencia de datos en el sector agropecuario. *Análisis de Datos Agropecuario*. Machala-Ecuador, Universidad Técnica de Machala, pp. 12-44.
- RODRÍGUEZ P.; PALOMINO N.; MONDACA, J. (2017). *El uso de datos masivos y sus técnicas analíticas para el diseño e implementación de políticas públicas en Latinoamérica y el Caribe*. Banco Interamericano de Desarrollo.
- ROJO, I. D. J. P.; SÁNCHEZ, A. A. C. (2019). Reinsurgencia de la etnografía en la era del Big Data: apuntes desde el sur global. *Virtualis*, v. 10, n. 19, pp. 42-56.
- RUVALCABA-GÓMEZ, E. A. (2020). Percepción del éxito sobre gobierno abierto en función del género: un análisis desde sociedad civil organizada y gobierno. *Revista iberoamericana de estudios municipales*, n. 21, pp. 61-87. Disponible en: https://www.scielo.cl/scielo.php?pid=S0719-17902020000100061&script=sci_arttext&tlng=pt. Acceso en: 12 nov 2021.
- SALGADO, D. (2017). Big Data en la Estadística Pública: retos ante los primeros pasos. *Economía industrial*, n. 405, pp. 121-129.
- SALVADOR, M.; RAMIÓ, C. (2020). Capacidades analíticas y gobernanza de datos en la Administración pública como paso previo a la introducción de la Inteligencia Artificial. *Revista del CLAD Reforma y Democracia*, n. 77, pp. 5-36.
- SERNA, M. S. (2021). Inteligencia artificial y gobernanza de datos en las administraciones públicas: reflexiones y evidencias para su desarrollo. *Gestión y Análisis de Políticas Públicas*, n. 26, pp. 20-32.
- SOSA ESCUDERO, W. (2020). *Big Data*. Ciudad Autónoma de Buenos Aires, Siglo XXI.
- STRAUSS, A.; CORBIN J. (1994). "Grounded theory methodology. An Overview". In: DENZIN, N. K.; LINCOLN, Y. S. (eds.). *Handbook of qualitative research*. Thousand Oaks, Sage.
- SUSAR, D.; AQUARO, V. (2019). Artificial Intelligence: opportunities and challenges for the public sector. In: 12TH INTERNATIONAL CONFERENCE ON THEORY AND PRACTICE OF ELECTRONIC GOVERNANCE. Melbourne, VIC, Australia, pp. 418-426. Disponible en: <https://dl.acm.org/doi/abs/10.1145/3326365.3326420>. Acceso en: 12 nov 2021.
- TANDOC JR, E. C.; LIM, Z. W.; LING, R. (2018). Defining "fake news" A typology of scholarly definitions. *Digital journalism*, v. 6, n. 2, pp. 137-153. Disponible en: <https://doi.org/10.1080/21670811.2017.1360143>. Acceso en: 16 abr 2022.

- TUFEKCI, Z. (2015). Algorithmic harms beyond Facebook and Google: emergent challenges of computational agency". *Colorado Technology Law Journal*, n. 13, pp. 203-218.
- VÁSQUEZ VALDIVIA, A. (2021). "Apertura y uso de datos para hacer frente al Covid-19 en América Latina. *Gestión Pública*, n. 88. Santiago, Comisión Económica para América Latina y el Caribe.
- VIGGO, H.; BYFUGLIEN, J.; JOHANNESSEN, R. (2003). Quality Issues at Statistics Norway. *Journal of Official Statistics*, v. 9, n. 3, pp. 287-303.
- WIRTZ, B. W.; WEYERER, J. C.; GEYER, C. (2019). Artificial Intelligence and the Public Sector – Applications and Challenges. *International Journal of Public Administration*, v. 42, n. 7, pp. 596-615. Disponible en: <https://doi.org/10.1080/01900692.2018.1498103>. Acceso en: 12 nov 2021.
- ZWITTER, A. (2014). Big Data ethics. *Big Data & Society*, v. 1, n. 2, pp. 1-6.

Texto recebido em 7/dez/2021
Texto aprovado em 9/mar/2022

