

VALORACIONES A LA CALIDAD DE LOS DATOS EN LAS ESTADÍSTICAS PÚBLICAS Y PRIVADAS, EN LA ERA DEL BIG DATA

ASSESSMENTS OF THE QUALITY OF THE DATA IN PUBLIC AND PRIVATE STATISTICS, IN THE ERA OF BIG DATA

Fernando Ariel Manzano

Universidad Nacional del Centro de la Provincia de Buenos Aires

fernandoarielmanzano@fch.unicen.edu.ar

<https://ordic.org/0000-0002-1513-4891>

Daniela Soledad Avalos

Universidad de Buenos Aires, Argentina

danielasoledad.av@gmail.com

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

Recibido: 21 de abril de 2022

Revisado: 14 de mayo de 2022

Aprobado: 12 de agosto de 2022

Cómo citar: Ariel Manzano, F. (2022). Valoraciones a la calidad de los datos en las estadísticas públicas y privadas, en la era del BIG DATA. *Bibliotecas. Anales de Investigación*;18(3), 1-12

RESUMEN

Objetivo: Describir las diferentes valoraciones respecto a la calidad de los datos en las estadísticas públicas. Se examinan los cambios producidos en la evaluación de la calidad en el marco de un aumento exponencial de la generación de los datos debido a las potencialidades del Big Data. **Diseño/ Metodología/ Enfoque:** El artículo es de carácter exploratorio. Mediante una amplia búsqueda de información especializada, se describe el conocimiento vinculado a esta problemática, desarrollada en los últimos años. **Resultados/ Discusión:** La estructura Big Data provienen del sector privado y no fueron generadas para propósitos estadísticos. La utilización en las Oficinas de estadísticas públicas implica lidiar diferentes fuentes de invalidez y múltiples causas que afectan la calidad de los datos. **Conclusiones:** Los institutos de estadística han ampliado el concepto de calidad y modificado sus principios de buenas prácticas con el propósito de incluir datos no generados con fines estadísticos. En la actualidad aún se carece de una posición unificada respecto a la viabilidad del uso de los macrodatos en las estadísticas nacionales. **Originalidad/ Valor:** El valor de las estadísticas oficiales se basa en la capacidad de generar información de calidad e imparcial para todos los agentes de la sociedad. Los avances tecnológicos de los últimos años, expresada en el *Big Data* y de sus disciplinas asociadas, sumado al aumento de la demanda de los usuarios de la apertura de datos, presentan nuevos retos en la evaluación de su calidad.

PALABRAS CLAVE: Calidad de los datos; Estadística pública; Big Data; Privacidad; Fuentes de datos.

ABSTRACT

Objective: Describe the different assessments of data quality in public statistics. It examines the changes in the assessment of quality in the context of an exponential increase in data generation due to the potential of Big Data. **Design/ Methodology/ Approach:** The article is exploratory in nature. By

means of an extensive search of specialised information, it describes the knowledge linked to this problem, developed in recent years. **Results/Discussion:** Big Data structures come from the private sector and were not generated for statistical purposes. The use in public statistical offices involves dealing with different sources of invalidity and multiple causes that affect the quality of the data. **Conclusions:** Statistical institutes have broadened the concept of quality and modified their principles of good practice in order to include data not generated for statistical purposes. At present, there is still no unified position on the feasibility of using big data in national statistics. **Originality/Value:** The value of official statistics is based on the ability to generate quality and unbiased information for all actors in society. The technological advances of recent years, expressed in Big Data and its associated disciplines, coupled with the increased user demand for open data, present new challenges in assessing their quality.

KEY WORDS: Data quality; Public statistics; Big Data; Privacy; Data sources.

INTRODUCCIÓN

La evolución de la gestión en las administraciones públicas se produce a partir de las transformaciones de las sociedades contemporáneas, especialmente, aquellas relacionadas con los cambios tecnológicos¹. Hacia finales del siglo XX en Europa y entrado el siglo XXI en América Latina, emergió el paradigma de la gobernanza² como modelo alternativo de gestión pública (Grandinetti y Nari, 2021). Este nuevo modo de entender al gobierno, distinto del jerárquico tradicional y de mercado, se caracteriza por mayores niveles de colaboración entre los gobiernos, las administraciones públicas y actores no gubernamentales en la confección de las políticas públicas (Zurbriggen, 2011; Grandinetti, 2013). En este contexto, los gobiernos deben responder a la demanda de la ciudadanía por mayor transparencia, legitimidad, acceso a la información y mecanismos efectivos de rendición de cuentas a la sociedad (Ruvalcaba-Gómez, 2020). Por otra parte, el valor de las estadísticas oficiales depende de su capacidad de crear conocimiento en la sociedad, y se las considera un bien público, al reunir las características de calidad, imparcialidad y posibilidad de ser difundidas de forma simultánea a todos los agentes (Ábrego Molina y Flores Mérida, 2021).

En los últimos años, a las fuentes y productores más tradicionales de datos e indicadores –como los censos de los sistemas nacionales de estadísticas oficiales, las encuestas y los registros públicos y privados–, se han añadido nuevas fuentes, potenciadas por las tecnologías de la información y de las comunicaciones (TIC). El uso intensivo de las TIC aumento la demanda de los usuarios de la apertura de datos para ser reutilizados, conduciendo al surgimiento de los datos públicos abiertos³ (DA) u *open data* –por su denominación en inglés– (Christodoulou *et al.*, 2018). La gestión de datos abiertos incrementó la complejidad de las administraciones públicas de los países⁴. Adicionando nuevas responsabilidades como velar por la calidad de los datos y su uso adecuado–cumpliendo con los estándares de calidad⁵, confidencialidad 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)..

Surge como gran desafío para las Estadísticas Oficiales, poder generar una metodología que otorgue precisión analítica sobre los grandes volúmenes de información, careciendo estos últimos de evidencia del fenómeno relevado (Bussi *et al.*, 2017; Luo *et al.*, 2019; Del-Fresno-García, 2014). El objetivo de este trabajo es clarificar la posición respecto a la viabilidad del uso de los macrodatos en los sistemas estadísticos nacionales. La relevancia de esta problemática, se expresa en los avances de *Big Data* y de sus

¹ La denominada Sociedad de la Información y el Conocimiento genero una redefinición histórica de las relaciones de poder, la productividad no depende del aumento de los factores de producción (capital, trabajo, recursos naturales), sino de la aplicación de conocimientos e información a la gestión, producción y distribución (Castells, 1997; Finkelievich *et al.*, 2011).

²A fines de la primera década del siglo XXI surgió el enfoque denominado gobernanza pública inteligente, que incorpora el impacto de los vertiginosos cambios tecnológicos. Aportando a la nueva manera de entender la política y la administración pública, basada en los valores emanados de una filosofía abierta y colaborativa (Castells, 2009; Subirats, 2012; Criado Grande, 2016).

³ La apertura de datos públicos contribuye a mejorar los niveles de transparencia (Concha y Naser, 2012).

⁴ Según la encuesta global de las Naciones Unidas sobre gobierno digital (Naciones Unidas, 2021), Argentina, Brasil, Chile, Colombia, Costa Rica, México y Uruguay se destacan dentro de la región por sus altos niveles de publicación de datos abiertos de gobierno (Naser *et al.*, 2017).

⁵ 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).

disciplinas asociadas, como machine learning, inteligencia artificial, ciencia de datos y otras relacionadas, generados fundamentalmente por el sector privado (Sosa Escudero, 2020 y 2021).

EVOLUCIÓN DE LA CALIDAD DE LAS ESTADÍSTICAS OFICIALES.

En la actualidad, el uso masivo de dispositivos móviles, el internet de las cosas, y la economía digital, conlleva que prácticamente toda actividad genere datos digitales. Los avances continuos en el intercambio de información entre los dispositivos que se manipulan diariamente y el incremento del número de usuarios, arroja en consecuencia grandes volúmenes de datos o Big Data⁶. 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 (Mayer-Schönberger y Cukier, 2013; Ábrego Molina y Flores Mérida, 2021). Una parte relevante del proceso de acumulación de grandes volúmenes de datos en los mercados, es conocer si la información obtenida es veraz. Resultando necesarios métodos cada vez más sofisticados para obtener un análisis de calidad de los datos recolectados. Así, el valor de la información, es lo que convierte al Big Data en un activo económico con demanda creciente por diversos actores del mercado⁷. La información posee determinados grados de utilidad según los usuarios, pero no todos los consumidores de datos están interesados en acceder a la misma información, por tanto, el valor de la misma varía según quién la demande (Labbé Figueroa, 2020).

Si bien no existe una definición de calidad de la información estadística acordada internacionalmente⁸. Tradicionalmente la calidad de la estadística en las instituciones públicas hacía referencia a la exactitud de las mismas (Viggo *et al.*, 2003)⁹. En tal sentido, algunas décadas atrás 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 el año 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 propios¹⁰–, que componen la calidad estadística (Naciones Unidas, 2004). A partir de estos principios, el concepto tradicional de calidad se fue ampliando¹¹, haciéndose gradualmente más comprensivo¹² y enfocándose particularmente en la relevancia, exactitud, la accesibilidad, puntualidad, comparabilidad, coherencia, completitud y la satisfacción de las expectativas de los diferentes tipos de usuarios (Elvers, 2002; Naciones Unidas, 2004).

El principio 5 de los Principios Fundamentales expresa que los datos para fines estadísticos pueden obtenerse de todo tipo de fuentes, pero no considera explícitamente la situación del *Big Data* (Salgado, 2017). Para abordar las nuevas fuentes de información, como el Big Data, se creó un Grupo de Trabajo Mundial de Naciones Unidas (United Nations Global Working Group on Big Data, 2016). En el año 2014 en Beijing este grupo llevo adelante la primera Conferencia Internacional sobre Big Data en las Estadísticas

⁶ Este término se ha acuñado en español como Macrodatos en algunas traducciones (Naciones Unidas, 2015). Los proyectos de macrodatos están conformados por agentes del sector privado y de institutos de investigación.

⁷ Debido a la introducción de *big analytics*, este grupo de información se ha transformado en un activo de valor muy relevante, que permite que existan empresas cuyos modelos de negocios, sus decisiones y estrategias, se encuentren basadas en el Big Data (Zúñiga, 2019, p. 209).

⁸ 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. Asimismo, identificaron tres aspectos principales de la calidad de la información estadística: pertinencia, exactitud y disponibilidad (Pettersson, 1992).

⁹ 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).

¹⁰ 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).

¹¹ Prácticamente se suelen brindar respuesta a todos los pedidos, salvo, aquellos casos que pongan en cuestión el secreto estadístico, e informando cuando corresponda las eventuales limitaciones de la información brindada (CEPAL, 2000). El concepto más amplio de calidad incluye criterios referidos a: relevancia, exactitud, oportunidad y puntualidad, accesibilidad y claridad, comparabilidad, coherencia y completitud (Naciones Unidas, 2004).

¹² 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).

Oficiales¹³. En el marco de este programa mundial destinado a generar recomendaciones para la utilización de datos de organizaciones privadas en las Estadísticas oficiales, se consensuó que labor principal se dirigiría hacia el marco de calidad de los macrodatos (Naciones Unidas, 2015)¹⁴. Así, el concepto ampliado de calidad en los sistemas estadísticos nacionales¹⁵ que incluye una gama amplia de fuentes de datos no recopilados con fines estadísticos –como los registros administrativos¹⁶, las fuentes Big Data, entre otras (Linerós, 2017)¹⁷–, es interpretado como un concepto multidimensional¹⁸, que contempla el grado de calidad, la oportunidad, el nivel de precisión, el costo y la carga que impondrán¹⁹ (Naciones Unidas, 2010; Lindenboim, 2011).

Diversas instituciones de carácter internacional han desarrollado guías para evaluar la calidad. Cabe mencionar, entre las más importantes, el Marco para la Evaluación de la Calidad de los Datos²⁰ (MECAD) del Fondo Monetario Internacional y el Código de las Buenas Prácticas de las Estadísticas Europeas de la Oficina de Estadísticas de las Comunidades Europeas (EUROSTAT) (Jiménez Mora, 2009).

En la etapa preparatoria de un operativo estadístico pueden acarrear conceptos no claros o erróneos. Así como también posteriormente, 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.

Durante el proceso de revisión los institutos de estadística deben tratar de conservar los originales en la mayor medida posible (CEPAL, 2011a; Naciones Unidas, 2011). Además, es responsabilidad informar a los usuarios a través de los metadatos²¹, 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 corregirlos²² (Giusti y Massé, 1997; CEPAL, 2020).

¹³ 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.

¹⁴ Siendo necesario recabar datos de organizaciones privadas como inputs, para producir estadísticas oficiales de calidad y relevantes (United Nations Global Working Group on Big Data, 2016).

¹⁵ 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).

¹⁶ 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.

¹⁷ Siendo necesario considerar todas las fuentes de datos de manera conjunta desde un enfoque multisectorial (Bussi *et al.*, 2017).

¹⁸ 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).

¹⁹ Los bajos resultados en la calidad de los censos de los años ochenta del siglo pasado en la región, 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).

²⁰ Este marco abarca todos los aspectos de la calidad de los datos y se organiza en torno a cinco dimensiones: garantías de integridad, rigor metodológico, exactitud y fiabilidad, utilidad de las estadísticas para el usuario y acceso a la información (Laliberté *et al.*, 2004).

²¹ La facilidad de interpretación de la calidad corresponde a la disponibilidad de información suplementaria y metadatos necesarios para interpretar la información estadística de manera apropiada. Esta información corresponde a tres categorías esenciales: a) Conceptos y clasificaciones en que se basan los datos. b) Metodología empleada para reunirlos y compilarlos. c) Medidas de exactitud de los datos (Echegoyen, 2004).

²² 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

A diferencia del tratamiento que se les da a las fuentes primarias, los registros administrativos como fuente de información requieren de un proceso de conversión estadística, centrado en el análisis de parámetros de calidad que, por la cantidad y diversidad de objetivos, variables, métodos de registro, tratamiento y archivo de los registros –ajenos a los fines estadísticos– resulta complejo (Gauna *et al.*, 2020). Pero en el caso de una arquitectura de Big Data²³, los institutos de estadísticas oficiales deben lidiar con un nivel de calidad de los datos aun inferior. Estas fuentes de información se caracterizan por su bajo nivel de estructuración de los datos²⁴ –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 datos tradicionales–. 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 en las estadísticas públicas.

El uso de las fuentes de Big Data por parte de los sistemas estadísticos de cada país, demanda una estrecha colaboración con el sector privado, para garantizar la confiabilidad de los datos (Bussi, 2017). Se debe tener presente que la estadística pública u oficial, es un instrumento estratégico para conocer objetivamente la realidad de la situación de un país y la generación de toma de decisiones²⁵ –como por ejemplo políticas públicas apropiadas–, basado en un modelo de calidad definido, documentado y conocido por los miembros del sistema (Tarazona, 2006; CEPAL, 2011b). La evaluación de la calidad de la estadística que se genera en los organismos públicos es considerada prioritaria.

EL BIG DATA Y LAS ESTADÍSTICAS OFICIALES.

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. 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)

El sector público debe lidiar con la baja calidad de las fuentes de información²⁶ de una arquitectura de Big Data (Pérez-Rave *et al.*, 2019), debido al bajo nivel de estructuración de los datos²⁷ –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–. Es responsabilidad de los institutos de estadística encontrar las diferentes fuentes de invalidez que afectan la calidad de los datos (Naciones Unidas, 2017). Pero, además, dado los principios éticos elementales del sector público, debe garantizar transparencia e identificar la existencia de errores²⁸ de estos grandes volúmenes de datos (Paliotta, 2018).

quedar registradas, así como también deben realizarse evaluaciones acerca de los efectos generados en la exactitud de los resultados (Naciones Unidas, 1980).

²³ El concepto “*Garbage in, garbage out*” utilizado en las ciencias de la información, refiere a que la entrada de datos sin sentido implica salida de información también sin sentido (Galimany Suriol, 2014). Se han añadido nuevas uves 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).

²⁴ Los datos según cierta tipología pueden diferenciarse en: estructurados (provenientes de bases de datos relacionales); semi-estructurados (principalmente archivos HTML, XML, y otros formatos similares); y no estructurados (incluye fotos, videos, audios y textos, entre otros) (Joyanes Aguilar, 2016). Existe cierto desconocimiento en la posibilidad de extraer información válida de los datos no estructurados.

²⁵ Por ejemplo, el Departamento Administrativo Nacional de Estadística de Colombia (DANE), implementó un Archivo Nacional de Datos (ANDA). El ANDA permite a los usuarios información relacionada con censos, encuestas por muestreo y uso estadístico de registros administrativos (Reíta Reyes y Salinas Hernández, 2016).

²⁶ El concepto “*Garbage in, garbage out*” utilizado en las ciencias de la información, refiere a que la entrada de datos sin sentido implica salida de información también sin sentido (Galimany Suriol, 2014). Se han añadido nuevas uves 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).

²⁷ Los datos según cierta tipología pueden diferenciarse en: estructurados (provenientes de bases de datos relacionales); semi-estructurados (principalmente archivos HTML, XML, y otros formatos similares); y no estructurados (incluye fotos, videos, audios y textos, entre otros) (Joyanes Aguilar, 2016). Existe cierto desconocimiento en la posibilidad de extraer información válida de los datos no estructurados.

²⁸ 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).

En el año 2014 se realizó en Beijing la primera Conferencia Internacional sobre Big Data en las Estadísticas Oficiales²⁹. 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 (Naciones Unidas, 2015).

El Grupo de Trabajo Mundial acordó que su labor se dirigiría más hacia el marco de calidad para los macrodatos, para lo cual agruparía los elementos de su programa de trabajo con las categorías de “calidad del aporte” (acceso a los datos) “calidad del procesamiento” (metodología de los macrodatos y métodos de estimación, el repositorio de proyectos sobre macrodatos y procesos de producción estadística) y “calidad de los productos” (aplicación de los macrodatos a las estadísticas oficiales, incluidos los indicadores de los Objetivos de Desarrollo Sostenible). En las sucesivas Conferencias se presentaron adelantos realizados en los proyectos de macrodatos con datos de telefonía móvil, de las redes sociales y satelitales para distintas aplicaciones estadísticas –estas implican alianzas con proveedores de datos que operan en todo el mundo–.

Diversas investigaciones remarcan que las nuevas fuentes de información que contiene una estructura Big Data, contienen múltiples dificultades, muchas de ellas vinculadas entre sí, como el acceso institucional a los datos – cambios legales en los países–, la metodología estadística para realizar inferencias respecto de las poblaciones de interés, y las nuevas tecnologías informáticas necesarias (Salgado, 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.

LA CALIDAD DE LAS FUENTES UTILIZADAS POR LAS ESTADÍSTICAS OFICIALES

Las administraciones públicas generan y recopilan grandes cantidades de datos, con diferentes finalidades y usos, distintas tipologías y grados de estructuración, asociadas a diferentes departamentos que, presentan importantes dificultades de integración (Jones, 2019; Kim, 2014). Los institutos de estadística deben realizar un análisis cuali-cuantitativo de los errores y las inconsistencias de los datos (Giusti y Massé, 1997; INDEC, 2020)– comprobaciones de coherencia a partir de especialistas temáticos (Naciones Unidas, 2011).–. Siendo unos de los objetivos establecer una metodología de control de la calidad³⁰, para detectar las posibles causas que afectan la calidad de los datos y orientar el modo de corregirlos (INDEC, 2020; Menéndez Campos y Zambrano, 2020). El proceso de corrección debe tratar de conservar la información original en la mayor medida posible (Naciones Unidas, 2011). En la selección de fuentes para fines estadísticos, se pondera la calidad, la oportunidad, el costo y la carga que impondrán³¹, en función de la magnitud de errores de diversa índole que posee la información (Lindenboim, 2011). Así como también, es necesario considerar las fuentes de datos de manera conjunta desde un enfoque multisectorial (Bussi *et al.*, 2017). A lo largo del proceso de producción de información los errores³² pueden introducirse en diferentes momentos–en la fase preparatoria, en el trabajo de campo y posterior al relevamiento– (Menéndez Campos y Zambrano, 2020). Se pueden distinguir tanto ventajas, como desventajas al utilizar las distintas fuentes tradicionales, ya sean: censos, encuestas estadísticas y registros administrativos³³.

En el inicio de un censo o encuesta se debe establecer la calidad en términos cuantitativos, mediante los umbrales o límites que pueden alcanzar los valores para considerarse correctos (CEPAL, 2011a; CELADE, 1999).

²⁹ La segunda Conferencia Internacional se realizó en 2015 en Abu Dabi (Emiratos Árabes Unidos), la tercera en 2016 en Dublín (Irlanda), la cuarta en 2017 en Bogotá (Colombia).

³⁰ 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).

³¹ La baja la calidad de los censos de los años ochenta en la región, echaron por tierra el convencimiento de mejora en el tiempo de los operativos, debido al adelanto tecnológico (Arretx y Chackiel, 1985). La encuesta sobre el estado de avance de los censos de 2020, arrojó que el 75 % realizaría cambios en sus controles, debido a que este factor se relaciona a la credibilidad y la confianza de la organización (CEPAL - CELADE, 2020).

³² 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 –suelen afectar a un grupo específico de variables o registros, y generan cierta distorsión en la distribución de los valores– (INDEC, 1987).

³³ Los Registros administrativos son datos relevados por autoridades nacionales y locales, para fines administrativos, de conformidad con fundamentos jurídicos diferentes de los de la legislación estadística (CEPAL, 2020, p. 10).

Dada la complejidad de realización de los censos de población estos presentan como desventaja, la alta probabilidad que se produzcan errores³⁴. Siendo relevante y valioso para los usuarios una correcta evaluación³⁵ de la magnitud de los mismos (Di Brienza *et al.*, 2003). Además, se debe limitar la cantidad de preguntas, para no generar un impacto negativo sobre la calidad de la información recolectada. Cabe destacar que el grado de calidad de los resultados finales del censo siempre son inferiores al de la información recolectada en terreno (CELADE, 1999). Respecto a las encuestas, su nivel de confiabilidad se encuentra asociado a los parámetros de calidad definidos en los diseños muestrales. Por último, los datos administrativos³⁶ presentan como principal inconveniente estar afectada por problemas de omisión, oportunidad y calidad de la información. Debido a que esta fuente ha sido diseñada para ser utilizada en la gestión.

Para el aprovechamiento y utilidad estadística de un registro administrativo es fundamental conocer el contexto de producción, análisis y utilización del registro. Los resultados generados deben mantener coherencia en el uso de clasificaciones y conceptos requeridos para un proceso estadístico de calidad³⁷ (CEPAL, 2011b). Para poder vincular la documentación de diferentes registros, resulta necesario contar con un número de identificación único para cada persona, hogar y vivienda. Asimismo, se requiere una armonización de los conceptos y definiciones que se utilizan en los distintos registros (Naciones Unidas, 2010).

Existe una fuerte demanda por mayor cantidad de preguntas en los censos, debido a las falencias en la cobertura y calidad de datos provisto por otras fuentes –encuestas o registros– (CEPAL, 2011a). Los productores de estadísticas oficiales deberán informar a los usuarios la calidad de los resultados estadísticos a través de metadatos (CEPAL, 2020). Así, los datos que el usuario utiliza pueden tener mayor o menor grado de integridad, precisión, validez y confiabilidad según, por un lado, cual sea el nivel y las características del error cometido y por otro, qué criterios se siguieron para tratar de corregirlos (Giusti y Massé, 1997).

CONCLUSIONES

La calidad es la piedra angular de la producción estadística oficial, entendida en sentido amplio. Esto implica que la calidad de los datos tenga una evaluación periódica por parte de los institutos de estadística nacionales conforme a los lineamientos internos y normas internacionales. Destacándose la veracidad de los datos como una dimensión fundamental de cara a la calidad de los resultados.

Se transita un periodo de aumento exponencial de la generación de datos digitales, así como también de mayor demanda de disponibilidad de datos por parte de los usuarios. Esto implica que los institutos de estadística nacionales, brinden información adicional como metadatos, pautas de estandarización e indicadores de calidad y cobertura, para garantizar una adecuada utilización de los datos.

A nivel internacional, el concepto de calidad se fue ampliando con el propósito de incluir datos no generados con fines estadísticos, como los datos administrativos y las fuentes de Big Data. Las potencialidades de este último, se encuentran sujeta a la generación de nuevos marcos de calidad y desarrollos de metodologías adecuadas, que permitan gestionar la baja calidad de esta información, así como también asegurar cuestiones relacionadas con la confidencialidad y el secreto estadístico.

Siendo la calidad el factor prioritario en la producción de datos en las estadísticas oficiales. Con independencia de los errores en los macrodatos, el desconocimiento de sus fuentes, su baja representatividad, entre otras debilidades. La posibilidad de brindar confiabilidad demanda un trabajo

³⁴ El censo presenta eventuales errores de omisión - el porcentaje de población que no fue censada (Arretx y Chackiel, 1985)-, debilitando su fortaleza de universalidad. Asimismo, la calidad de la información puede ser deficiente al ser obtenida mediante empadronadores no especializados.

³⁵ El censo presenta eventuales errores en las magnitudes –por omisión o por duplicación-, en el recuento de las viviendas, hogares y /o individuos (errores de cobertura); errores de los encuestadores; registros tardíos (CEPAL, 2011a; Naciones Unidas, 2011), entre otras cuestiones ligadas a la integridad y validez de la información relevada (errores de calidad) (Giusti y Massé, 1997; Arretx y Chackiel, 1985). Una medida de la calidad de los datos en término de análisis de contenido se detecta a partir del nivel de no respuesta, multimarca o inconsistencia de los datos, de acuerdo con pautas de validación predeterminadas (INDEC, 2020).

³⁶ En América Latina, los registros de hechos vitales empezaron con los registros parroquiales. Actualmente, la mayoría de países de la región cuenta con una institución de Registro Civil. No obstante, las estadísticas vitales suelen presentar problemas de cobertura y calidad de la información, además de registro tardío, sobrerregistro, interrupciones en el tiempo, etc. (Cavenaghi, 2012).

³⁷ La pérdida de calidad puede deberse a: cambios en las clasificaciones utilizadas, uso de nomenclaturas y variables de clasificación que no permiten relacionar la información con otras fuentes, alteración de su regularidad por cambios sistemáticos, entre otras (Hermida, 2016).

conjunto con el sector privado. En este marco, se destaca la ausencia de una posición unificada respecto a la viabilidad del uso de los macrodatos en los institutos de estadísticas nacionales.

REFERENCIAS BIBLIOGRÁFICAS

Ábrego Molina, Víctor Hugo y Flores Mérida, Antony (2021). Datificación crítica: práctica y producción de conocimiento a contracorriente de la gubernamentalidad algorítmica. Dos ejemplos en el caso R

REFERENCIAS BIBLIOGRÁFICAS

Ábrego Molina, Víctor Hugo y Flores Mérida, Antony (2021). Datificación crítica: práctica y producción de conocimiento a contracorriente de la gubernamentalidad algorítmica. Dos ejemplos en el caso mexicano. *Administración Pública y Sociedad (APyS)*, (11), 211-231.
<https://revistas.unc.edu.ar/index.php/APyS/article/view/33025>

Arretx, Carmen, y Chackiel, Juan (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*.
<https://repositorio.cepal.org/handle/11362/12712>

Bussi, Javier; Marí, Gonzalo Pablo y Méndez, Fernanda (2017). El desafío del big data en estadísticas oficiales en Argentina. Vigesimosegundas Jornadas "Investigaciones en la Facultad" de Ciencias Económicas y Estadística. Noviembre de 2017. Universidad Nacional de Rosario.
<https://rephip.unr.edu.ar/handle/2133/10331>

Castells, Manuel (2009). *Communication power*. New York, NY: Oxford University Press.

Castells, Manuel (1997) *La era de la información. Economía, sociedad y cultura*. Vol. 1 *La sociedad red*, Madrid: Alianza. <http://200.2.15.132/handle/123456789/32960>

CavenaghI, Susana (2012). Estimaciones y proyecciones de población en América Latina: desafíos de una agenda pendiente. Asociación Latinoamericana de Población. Rio de Janeiro (Brasil).
<http://docplayer.es/75504819-Estimaciones-y-proyecciones-de-poblacion-en-america-latina-desafios-de-una-agenda-pendiente.html>

CELADE (1999). América Latina: aspectos conceptuales de los censos del 2000. (Seminario Censos 2000: diseño conceptual y temas a investigar en América Latina). Santiago de Chile. Recuperado de https://repositorio.cepal.org/bitstream/handle/11362/5548/S9950308_es.pdf

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. Recuperado de https://celade.cepal.org/censosinfo/documentos/INFORME_CENSOS_RONDA_2020_LAyCARIBE.pdf

CEPAL (2020). *Ley Genérica sobre Estadísticas Oficiales para América Latina (LC/CEA.10/8)*. Santiago.
https://repositorio.cepal.org/bitstream/handle/11362/45253/1/S2000046_es.pdf

CEPAL (2011a). Guía para asegurar la calidad de los datos censales. Serie N°74.
https://repositorio.cepal.org/bitstream/handle/11362/5515/S1100929_es.pdf?sequence=1&isAllowed=y

CEPAL (2011b). Código Regional de Buenas Prácticas en Estadísticas para América Latina y el Caribe. Sexta reunión de la CEA-CEPAL, celebrada en Bávaro, República Dominicana en noviembre de 2011.
<https://www.parlamericas.org/uploads/documents/CEPAL%20Resources%20-%20codigo-regional-buenas-practicasALC.pdf>

CEPAL (2000). Sexto Taller Regional sobre Indicadores sobre el Desarrollo Social. Buenos Aires 15 a17 noviembre 2000. <http://hdl.handle.net/11362/20000>

Christodoulou, Paraskevi; Decker, Stefan; Douka, Aikaterini Vasiliki; Komopoulou, Charalampia; Peristeras, Vasilios; Sgagia, Sofia; Tsarapatsanis, Vaios; y Vardouniotis, Dimosthenis (2018). Data Make the Public Sector Go Round. En P. Parycek, O. Glassey, M. Janssen, H. J. Scholl, E. Tambouris, E. Kalampokis, S. Virkar (eds.). *Electronic Government. EGOV 2018. Lecture Notes in Computer Science* (vol. 11.020). Springer. https://doi.org/10.1007/978-3-319-98690-6_19

- Concha, Gaston y Naser, Alejandra (2012), "Datos abiertos: un nuevo desafío para los gobiernos de la región", serie Gestión Pública, N° 74 (LC/IP/L.313), Santiago, Comisión Económica para América Latina (CEPAL), marzo. Recuperado de <https://repositorio.cepal.org/handle/11362/7331>
- Criado Grande, Ignacio (2016). Las administraciones públicas en la era del gobierno abierto. Gobernanza inteligente para un cambio de paradigma en la gestión pública. *Revista de Estudios Políticos*, 173, 245–275. <http://dx.doi.org/10.18042/cepc/rep.173.07>
- Del-Fresno-García, Miguel (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*, 23(3), 246–252. <https://doi.org/10.3145/epi.2014.may.04>
- Di Brienza, María; Freitez, Anitza y León, Armando (2003). ¿El censo 2001, revela avances en la calidad de la información? *revista Temas de Coyuntura*, (47). https://www.academia.edu/43335466/El_censo_2001_revela_avances_en_la_calidad_de_la_informaci%C3%B3n
- Dufty, David; Bérard, Helene; Lefranc, Sylvie y Signore, Marina (2014). "A suggested framework for the quality of big data", *UNECE big data Qual. task team*. <https://statswiki.unece.org/download/attachments/108102944/Big%20Data%20Quality%20Framework%20-%20final-%20Jan08-2015.pdf?version=1&modificationDate=1420725063663&api=v2>
- Echegoyen, G. (2004). Registros administrativos, calidad de los datos y credibilidad pública: presentación y debate de los temas sustantivos de la segunda reunión de la Conferencia Estadística de las Américas de la CEPAL (Vol. 21). United Nations Publications. <https://www.cepal.org/es/publicaciones/4725-registros-administrativos-calidad-datos-credibilidad-publica-presentacion-debate>
- Elvers, Eva (2002). Comparison of Survey and Register Statistics. The International Conference on Improving Surveys, agosto 2002, Denmark, University of Copenhagen
- Finquelievich, Susana; Prince, Alejandro y Jolías, Lucas (2011): "Territorios digitales: el camino a la Sociedad de la Innovación. El caso de San Luis, Argentina", en: *Revista Iberoamericana de Estudios Municipales*, RIEM – ICHEM, Año 2, N° 3, Primer Semestre, marzo de 2011, pp. 41-62.
- Galimany Suriol, Aleix (2014). La creación de valor en las empresas a través del Big Data. Grado de Administración y Dirección de Empresas. Universidad de Barcelona. <http://diposit.ub.edu/dspace/handle/2445/67546>
- Gauna, N., Roggi, C., y Zuloaga, N. (2020). Los registros administrativos en la construcción y consolidación del Sistema Estadístico de la Ciudad. *Población de Buenos Aires*, 17(29). <https://www.redalyc.org/journal/740/74065319007/html/>
- Giusti, Alejandro y Massé, Gladys (1997). Aspectos conceptuales relativos a la evaluación de calidad. En INDEC, *Evaluación de la calidad de datos y avances metodológicos Serie J n° 2*. Buenos Aires.
- Grandinetti, Rita Maria y Nari, Patricia Olga (2021). Ciudades latinoamericanas: la necesidad de ser capaces de gestionar una nueva agenda urbana. *A&P Continuidad*, 8(14). doi: <https://doi.org/10.35305/23626097v8i14.303>
- Grandinetti, Rita (2013). Tecnología de La Administración Pública. <https://tecnoadministracionpub.files.wordpress.com/2014/08/1grandinettinotas-sobre-los-aportes-de-la-gobernanza-y-la-nueva-institucionalidad.pdf>
- Hermida, Mariano (2016). Metodologías para el tratamiento estadístico de los registros administrativos. In *V Encuentro Latinoamericano de Metodología de las Ciencias Sociales (Mendoza, 16 al 18 de noviembre de 2016)*. <http://sedici.unlp.edu.ar/handle/10915/108785>
- Instituto Nacional de Estadística y Censos (INDEC) (2020). Informe de evaluación de calidad cuantitativa de la segunda prueba piloto: hacia la ronda 2020. (Documentos de trabajo; 33). Ciudad Autónoma de Buenos Aires.

- https://www.indec.gov.ar/ftp/cuadros/poblacion/informe_IIPP_calidad_cuantitativa_censo_ronda2020.pdf
- Instituto Nacional de Estadística Y Censos (INDEC) (1987). III Curso de informática para estadísticos. Modulo 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.
- Jiménez Mora, José Manuel (2009). Gestión de la calidad de la estadística pública. Tesis Doctoral. Universidad Veracruzana. <https://148.226.24.32/handle/123456789/47679>
- Jones, Matthew (2019). Whatwetalkaboutwhenwetalkabout (big) data. *TheJournal of StrategicInformationSystems*,28(1), 3-16. <https://doi.org/10.1016/j.jsis.2018.10.005>
- Joyanes Aguilar, Luis (2016). *Big Data, Análisis de grandes volúmenes de datos en organizaciones. Primera edición.*
- Kim, Gang Hoon; Trimi, Silvana y Chung, Ji Hyong (2014). Big Data Applications in theGovernment Sector: A ComparativeAnalysisamongLeadingCountries. *Communicationsofthe ACM*, 57(3), 78-85. <https://doi.org/10.1145/2500873>
- Labbé Figueroa, M. F. (2020). Big Data: Nuevos desafíos en materia de libre competencia. *Revista chilena de derecho y tecnología*, 9(1), 33-62. https://www.scielo.cl/scielo.php?script=sci_arttext&pid=S0719-25842020000100033
- Laliberté, Lucie; Grünewald, Werner y Probst, Laurent (2004). Data quality: a comparison of IMF's Data Quality Assessment Framework (DQAF) and EUROSTAT's quality definition. International Monetary Fund. Washington, D.C. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.182.5048&rep=rep1&type=pdf>
- Lindenboim, Javier (2011). Las estadísticas oficiales en Argentina ¿Herramientas u obstáculos para las ciencias sociales? Trabajo y Sociedad. N° 16, vol. XV, V. <https://dialnet.unirioja.es/servlet/articulo?codigo=3731214>
- Lineros, Eduardo Martin (2017). El trinomio dato-información-conocimiento. In *Manual sobre utilidades del big data para bienes públicos* (pp. 35-48). Edimema.
- Luo, Jar Der; Liu, Jifan; Yang, Kunhao y Fu, Xiaoming (2019). Big data researchguidedbysociologicaltheory: A triadicdialogueamongbig data analysis, theory, and predictivemodels. *TheJournal of ChineseSociology*,6(1), 11. <https://doi.org/10.1186/s40711-019-0102-4>
- Matus C. (2007). Dimensiones de la calidad según OECD y EUROSTAT. Instituto Nacional de Estadísticas. Santiago, Chile.
- Mayer-Schönberger, V., y Cukier, K. (2013). Big data: la revolución de los datos masivos. Turner.
- Menéndez Campos, Jesica y Zambrano, María (2020). Calidad de los censos tradicionales de población y vivienda: Evaluación de sus etapas. *Matemática*, 18(2). <http://www.revistas.espol.edu.ec/index.php/matematica/article/view/752>
- Molina, V. H. Á., y 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 (APyS)*, (11), 211-231. <https://revistas.unc.edu.ar/index.php/APyS/article/view/33025>
- Monleon-Getino, Antonia (2015). El Impacto del Big Data en la Sociedad de la Información. Significado y utilidad. *Historia Y Comunicación Social*, 20(2), 427-445. http://dx.doi.org/10.5209/rev_HICS.2015.v20.n2.51392
- Naciones Unidas (2021). “Objetivos de Desarrollo Sostenible”. <https://www.un.org/sustainabledevelopment/es/>
- Naciones Unidas (2017). Principles and Recommendations for Population and Housing Censuses. Revisión 3. Disponible en: <https://unstats.un.org/unsd/demographic-social/Standards-and->

- Naciones Unidas (2015). Informe del Grupo de Trabajo Mundial sobre los Macrodatos en las Estadísticas Oficiales. E/CN.3/2016/1. <https://undocs.org/pdf?symbol=es/E/CN.3/2016/1>
- Naciones Unidas (2011). Manual de revisión de datos de los censos de población y vivienda. Revisión 1. Departamento de Asuntos Económicos y Sociales. *División de Estadística*. https://unstats.un.org/unsd/publication/SeriesF/seriesf_82rev1s.pdf
- Naciones Unidas (2010). United Nations E-Government Survey 2010: Leveraging e-government at a time of financial and economic crisis.
- Naciones Unidas (2004). Manual de Organización Estadística. Nueva York, EEUU.
- Naciones Unidas (1980). Principios y recomendaciones para la elaboración de Censos de Población y habitación. Serie M N°67. New York.
- Naser, Alejandra; Ramírez-Alujas, Álvaro y Rosales, Daniela (2017). *Desde el gobierno abierto al Estado abierto en América Latina y el Caribe*. Cepal. https://repositorio.cepal.org/bitstream/handle/11362/44769/1/S1601154_es.pdf
- Ontiveros, Emilio y López Sabater, Verónica (2017). Economía de los datos. Ariel. <https://www.fundacioncarolina.es/wp-content/uploads/2018/11/Libro-Economia-de-los-Datos-Ontiveros.pdf>
- Paliotta, Achille Pierre (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 (CESOT)*, (10), 63-94. <http://ojs.econ.uba.ar/index.php/CESOT/article/view/1405>
- Pérez-Rave, Jorge; Correa Morales, Juan Carlos y González Echavarría, Fabián (2019). Metodología para explorar datos abiertos de accidentalidad vial usando Ciencia de Datos: Caso Medellín. *Ingeniare. Revista chilena de ingeniería*, 27(3), 495-509. https://scielo.conicyt.cl/scielo.php?pid=S0718-33052019000300495&script=sci_arttext
- Pettersson, Hans. (1992). Control de Calidad en Estadísticas Procedentes de Registros y Archivos Administrativos. Instituto Vasco de Estadística. Euskadi.
- Reíta Reyes, Jorge Eduardo y Salinas Hernández, Héctor Javier (2016). Análisis de la viabilidad de la implementación de redes Big Data en Colombia. Universidad Distrital FJDC Facultad Tecnológica.
- Rodríguez, Patricio; Palomino, Norma y Mondaca, Javier (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. <https://publications.iadb.org/publications/spanish/document/El-uso-de-datos-masivos-y-sus-tecnicas-analiticas-para-el-diseno-e-implementacion-de-politicas-publici.pdf>
- Ruvalcaba-Gómez, Edgar Alejandro (2020). Percepción del éxito sobre Gobierno Abierto en función del Género: un análisis desde Sociedad Civil Organizada y Gobierno. *Revista iberoamericana de estudios municipales*, (21), 61-87. https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0719-17902020000100061&lang=es
- Salgado, David (2017). Big Data en la Estadística Pública: retos ante los primeros pasos. *Economía industrial*, (405), 121-129. <https://dialnet.unirioja.es/servlet/articulo?codigo=6207520>
- Salvador, Miquel y Ramió, Carles (2020). Capacidades analíticas y gobernanza de datos en la Administración pública como paso previo a la introducción de la Inteligencia Artificial. *Reforma Democr. Rev. CLAD*, 77, 5-36. <https://clad.org/wp-content/uploads/2021/04/077-01-SR.pdf>
- Sosa Escudero, Walter (2021) Big data y ciencia de datos: conceptos, oportunidades y desafíos, Nota informativa, Organización Internacional del Trabajo. https://www.ilo.org/wcmsp5/groups/public/---americas/---ro-lima/---srosantiago/documents/publication/wcms_769307.pdf

Sosa Escudero, Walter (2020). Big data: desafíos para la política pública. *Revista del CLAD Reforma y Democracia*, (76), 57-70. <https://clad.org/wp-content/uploads/2021/01/076-03-SEscudero.pdf>

Subirats, Joan (2012), “Nuevos tiempos, ¿nuevas políticas públicas? Explorando caminos de respuesta”, *Reforma y Democracia*, N° 54, Caracas. <http://www.clad.org/portal/publicaciones-del-clad/revista-clad/reformademocracia/articulos/054-octubre-2012/Subirats.pdf>.

Tarazona, Jorge. (2006). Proyecto certificación de calidad de la información básica. *revista de la información básica*, 6 (1), 28-35.

United Nations Global Working Group on Big Data (2016). *Recommendations for access to data from private organizations for Official Statistics*. Dublin: United Nations.

Vásquez Valdivia, Alvaro (2021). “Apertura y uso de datos para hacer frente al COVID-19 en América Latina”, serie *Gestión Pública*, N° 88 (LC/TS.2021/98), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL). <https://repositorio.cepal.org/handle/11362/47172>

Viggo, Hans; Byfuglien, Jan y Johannessen, Randi (2003). Quality Issues at Statistics Norway. *Journal of Official Statistics*, 9 (3), 287-303. <https://ssb.brage.unit.no/ssb-xmlui/handle/11250/178164>

Zúñiga, G. (2019). "Big data y los desafíos que plantea al abuso de posición de dominio". *Revista de Actualidad Mercantil*. (6), 208-226. <https://bit.ly/30m3xzC>

Zurbriggen, Cristina (2011). Gobernanza: una mirada desde América Latina. *Revista Perfiles Latinoamericanos - Flacso*, 38, 39–64. http://www.scielo.org.mx/scielo.php?pid=S0188-76532011000200002&script=sci_arttext