Organización Internacional del Trabajo

Nota de la OIT

Nota informativa

Febrero de 2021

Big data y ciencia de datos: conceptos, oportunidades y desafíos[1]

Walter Sosa Escudero[2]

Principales puntos

- El mundo del trabajo y las instituciones laborales han comenzado en forma creciente a utilizar *big data*. Esta nota discute el ámbito de acción de varios conceptos relacionados con este fenómeno, entendido como una combinación de datos masivos con algoritmos, fórmulas y métodos computacionales y matemáticos.
- La nota define varios conceptos asociados, discute algunos casos concretos, especifica las ventajas y limitaciones del fenómeno y ejemplifica con casos concretos la utilidad de *big data* en la política pública.

El fenómeno de *big data* parece invadir todos los aspectos de la vida cotidiana, profesional y académica. Como sucede con todas las tecnologías innovadoras, la sociedad parece radicalizarse ante la irrupción de un fenómeno disruptivo, y *big data* no es la excepción.

Están los que opinan que los datos y algoritmos serán capaces de desplazar al método científico, como dice Chris Anderson (2008) en su polémico artículo en la revista Wired y quienes solo ven una moda pasajera, como en la enérgica charla de Tim Harford (2014).

Tal vez lo más interesante y relevante sea adoptar una postura prudente y sanamente dubitativa, capaz de sopesar ventajas y limitaciones de lo que se puede esperar de este fenómeno masivo. Y este es el objetivo y la postura de esta nota técnica: brindar un panorama abarcativo del ámbito de *big data*, incluyendo sus disciplinas asociadas, *como machine learning*, ciencia de datos y su relación con la estadística, la computación y la matemática.

La nota comienza relatando algunos casos relevantes del uso de datos masivos y algoritmos en el ámbito público. Recién ahí procede a aislar con precisión el fenómeno de *big data* y sus disciplinas afines. Luego, discute algunas limitaciones que, entendidas como desafíos, hablan del enorme potencial que tienen los datos en la cosa pública.

El artículo cierra con un breve relato de una experiencia concreta del uso de *big data* en el sector público.

[1] El autor agradece los valiosos comentarios de Roxana Maurizio, Marcela Cabezas y Fabio Bertranou.

[2] Profesor plenario de la Universidad de San Andrés. Investigador principal, CONICET (Argentina).

Tres casos y una definición

Teléfonos celulares para medir la pobreza en Ruanda

Ruanda es un país extremo en lo que se refiere a pobreza estructural, no hay que argumentar demasiado a fines de establecer la urgencia y la relevancia de medir correctamente la pobreza. Las mismas razones que explican este fenómeno dan cuenta de la imposibilidad práctica de confiar en un sistema de instituciones y encuestas, como las disponibles en los países de desarrollo intermedio y avanzado. A modo de ejemplo, el popular enfoque de líneas (ver Gasparini, Cicowiez y Sosa Escudero, 2013, para más detalles) demanda medir sistemáticamente los ingresos de las personas y el valor de una canasta de bienes y servicios que se entiende debe poder comprar una familia para no ser pobre.

Blumenstock et al. (2015) notan que en Ruanda existe una relación relevante entre la intensidad del uso de teléfonos celulares y el bienestar de la población. Sobre la base de una pequeña encuesta de bienestar y datos de uso de teléfonos, entrenan un modelo que les permite predecir el bienestar y, posteriormente, la pobreza.

No exento de críticas, el ejercicio es replicable en varias regiones del país y en distintos periodos, lo que permite cierta comparabilidad regional y temporal. Se trata de un primer ejemplo relevante de cómo datos ya disponibles (como el uso de celulares) pueden ser explotado para enfrentar una cuestión urgente, que en su tratamiento tradicional (en base a encuestas científicas) demanda un enorme esfuerzo institucional, extremadamente costoso para un país como Ruanda.

Precios e inflación en Argentina

A partir del 2007 hubo serias dudas acerca de la forma en la que se medía oficialmente la inflación en Argentina, coincidiendo con la intervención del Instituto Nacional de Estadística y Censos (INDEC), problema particularmente relevante en un país con una larga historia de crecimiento sostenido del nivel de precios. Sin entrar en detalles, un índice de precios requiere relevar sistemáticamente un conjunto de precios, que luego son agregados en un índice que los pondera –de acuerdo al peso que cada uno tiene dentro del total de gasto- en base a las cantidades consumidas por el conjunto de los hogares. Consecuentemente, la construcción de un índice de precios requiere un complejo y costoso sistema de encuestas de precios y cantidades.

A fines de evaluar la efectividad de las mediciones oficiales de la inflación, Cavallo y Rigobon (2016) reconstruyeron el índice de precios argentino en base a información disponible en internet. Más concretamente, los autores construyeron una base de datos de precios en base a técnicas de web scrapping, que “chupan” información de precios, cantidades y calidades de páginas web comerciales. Luego de una delicada tarea de “curación” de precios, los autores reportan un índice que guarda notorias similitudes con el construido sobre la base de encuestas estándar.

Este tipo de índice permite no solo monitorear la relevancia del índice oficial sino también focalizar en ciertas canastas alternativas, o ser medido con una frecuencia mucho más alta que la oficial, lo cual es crucial en tiempos de aceleración de la inflación, o ante la aparición de fenómenos relevantes, como la epidemia de coronavirus.

Impuestos a las ventas en los Estados Unidos

Einav et al. (2014) usan una inteligente estrategia para medir el impacto del impuesto a las ventas sobre la conducta de los consumidores. Un experimento ideal requeriría alterar el impuesto en una región o período, y dejando cualquier otro factor alternativo constante, sopesar el efecto del impuesto sobre las ventas, confiando en que ningún otro causante de la conducta se haya alterado. Alternativamente, el trabajo en cuestión tuvo acceso a la base de la empresa de comercio online Ebay, que guarda registro de la secuencia de *clicks* que conducen (o no) a una venta final. Concretamente, un usuario de Ebay pasa cierto tiempo buscando un producto, recorriendo alternativas de precios, cantidades y calidades. Luego de cierta búsqueda

decide proceder a la compra, tras lo cual se le revela el costo de envío y el impuesto a las ventas.

La disponibilidad de datos sobre todo el proceso de compra permite a estos investigadores aislar el impacto que tienen los precios, las calidades, el costo de envío del que tiene el impuesto a las ventas. Es decir, la sucesión de *clicks* que sigue un comprador funciona como una suerte de "experimento natural" que permite aislar el impacto del impuesto a las ventas de cualquier otro factor concurrente.

Los casos, en perspectiva

Estos tres casos -los datos de uso de celulares en Ruanda, los precios online en el caso de Argentina, y la secuencia de *clicks* en el caso de EEUU- tienen en común un crucial elemento: la principal fuente de datos para su ejecución se hallaba disponible antes de su implementación, en formato "electrónico", nunca generada por ninguna encuesta ni acción explícita.

También tienen en común el hecho de que los tres sustituyen o complementan mecanismos tradicionales para enfrentar cuestiones cruciales de la operatoria del sector público, como la medición de la pobreza, la inflación o el impacto de la política económica.

Adicionalmente, los tres ejemplos demandan algún esfuerzo estadístico y computacional, ya sea para extraer o sistematizar los datos ya disponibles, o para construir modelos que permitan medir, caracterizar o predecir una variable concreta, como la pobreza, la inflación o el impacto de alterar los impuestos.

Los casos difieren en su uso. El primero, es un ejercicio de predicción condicional, es decir, de predecir una variable (la pobreza) en base a otra (la intensidad de uso de celulares). El segundo caso (inflación en Argentina) es un ejercicio de medición, que simplemente intenta proveer información donde o no la hay o hay serias dudas de la ya disponible. El tercero (impuesto a las ventas) es de una naturaleza completamente distinta ya que es un ejercicio de evaluación de impacto, que intenta cuantificar el canal causal que va de los impuestos a la conducta de los consumidores.

¿De qué hablamos cuando hablamos de Big Data?

Los casos discutidos sugieren dos elementos clave en su ejecución: la naturaleza concreta de ciertos datos y la intervención de procedimientos computacionales - estadísticos.

En relación a los datos, los tres casos se basan en datos que no surgen de ninguna encuesta ni experimento diseñado a ningún fin específico, sino de información proveniente del hecho de que las personas interactúan espontáneamente con objetos (reales o no) interconectados, como los teléfonos celulares en el caso de Ruanda, o el comercio electrónico online como en el caso de la inflación en Argentina o el impuesto a las ventas en EEUU.

> La naturaleza espontánea de *big data* es posiblemente su característica más saliente, la que explica el copioso volumen del fenómeno y su prevalencia.

Esta espontaneidad es clave en este nuevo paradigma de datos: la existencia del dato es una mera consecuencia de ver cómo interactúan las personas o instituciones, sin que medie un esfuerzo explícito (en forma de una encuesta, experimento o ejercicio administrativo) para que ello ocurra. Es esta misma espontaneidad lo que explica la masividad del fenómeno, su velocidad de ocurrencia y su fuerte heterogeneidad.

La contracara de los datos son los métodos matemáticos, computacionales, estadísticos que se usan para procesar y sistematizar estos datos espontáneos. Ciertamente, la estadística tradicional cumple un rol crucial en este fenómeno. La idea de *machine learning* es, tal vez, una visión que integra a la estadística y a la computación, en donde los modelos se construyen (más que se estiman) dentro del proceso de análisis, en forma iterativa, a la luz de nueva información disponible y de la performance del modelo. Más allá de ser una terminología de moda, la ciencia de datos, sugiere un enfoque integrador e interactivo de las actividades que

involucran al análisis moderno de datos, que convoca a la estadística, a la matemática, a la computación, pero también a las disciplinas "de dominio" (desde la economía a la antropología, pasando por las ciencias naturales, de la gestión y el comportamiento) y también a la comunicación y al diseño. Es tal vez esta visión "panóptica" lo que hace que la ciencia de datos exista más allá de las disciplinas tradicionales que integra.

En este contexto es que *big data* puede ser entendido como un fenómeno de datos masivos que surgen de la interacción espontánea con dispositivos interconectados, que son estudiados y procesados por métodos computacionales y numéricos.

Cuatro desafíos

La discusión anterior sugiere que la naturaleza espontánea de *big data* es posiblemente su característica más saliente, la que explica el copioso volumen del fenómeno y su prevalencia. Los cuatro casos sugieren el enorme potencial del fenómeno, capaz de complementar o directamente sustituir estrategias tradicionales como las encuestas, los experimentos y los censos.

Pero es la misma espontaneidad lo que establece los desafíos del fenómeno de *big data*. Los datos tradicionales, de una encuesta o experimento, tiene una estructura clara, es decir, un mecanismo, formal o no, que establecen una relación explícita entre el dato y aquello que pretende representar, como el diseño muestral en el primer caso, o experimental en el segundo.

> Los datos de *big data*, al ser espontáneos, no tienen estructura obvia que permita asociarlos directamente con una población o fenómeno de referencia. Entonces, tal vez el desafío más importante que enfrenta *big data* es construir esta estructura, que permita hacer un uso efectivo de los datos (…)

Por el contrario, los datos de *big data*, al ser espontáneos no tienen ninguna estructura obvia que permita asociarlos directamente con una población o fenómeno de referencia. Entonces, tal vez el desafío más importante que enfrenta *big data* es construir esta estructura, que permita hacer un uso efectivo de los datos, más allá de su volumen y la velocidad a la cual se encuentran disponibles. Esta sección discute algunos desafíos concretos en relación a esta falta de estructura. Ciertamente, no se trata de minimizar el impacto crucial de *big data* sino todo lo contrario: cuando estas limitaciones son entendidas como auténticos desafíos, una vez sorteadas, el potencial de *big data* y algoritmos puede ser enorme, en particular en el ámbito público y de la política social. Esta sección se basa en Sosa Escudero (2020).

Big data no es (ni será) todos los datos

El flamante premio Nobel a Esther Duflo, Abhijit Banerjee y Michael Kremer, en 2019, es un indicio claro de que una parte fundamental del análisis empírico de la política requiere experimentos correctamente diseñados, como los de la agronomía o la medicina. La esencia del experimento se basa en comparar hechos con sus contrafácticos. A modo de ejemplo, a fines de evaluar si una droga tiene un efecto (causal) es relevante comparar a una persona a quien se la administró dicha droga con exactamente la misma persona en la circunstancia en la que no se la administró. Un serio problema filosófico es que la decisión de administrar una droga automáticamente impide poder observar a la misma persona en su circunstancia contrafáctica, es decir la misma persona no habiéndosele administrado la droga. Es decir, es posible observar sus acciones, pero jamás sus contrafácticos. El experimento científico es un hito en la historia de la ciencia, entendido como una herramienta para la construcción de contrafácticos. En particular, un experimento científico asigna al azar una droga a un grupo y un placebo a otro, de modo que lo único que, a fines de los objetivos del experimento, distingue a un grupo y al otro es si fue administrada la droga. Desde esta perspectiva, se entiende que una persona en el grupo de tratamiento es como si fuese la misma persona del grupo de control pero que fue tratada

con la droga. Este es el sentido en el que un experimento no observa, sino que construye información contrafáctica.

Por definición, los datos de *big data* son observacionales, de modo que no dicen nada per se acerca de la información contrafáctica. Entonces, *big data* no puede ser todos los datos, y, siendo optimistas tal vez sea capaz de revelar, a lo sumo, la mitad de los datos, producto de las acciones, pero, por definición, no de sus contrafácticos. Esta discusión es bastante más que un ejercicio intelectual-filosófico: la evaluación de políticas constituye un eje fundamental de la visión moderna de la cuestión pública, como lo reafirma el flamante premio Nobel asignado a esta reciente y crucial aproximación empírica a través del análisis científico de los resultados de una política.

Un importante punto es que si bien *big data* per se no revela información contrafáctica, es una importantísima fuente para la construcción de datos cuasi experimentales. El estudio de Einav et al. (2014) discutido anteriormente, sobre los efectos causales del impuesto a las ventas, es una gran ilustración de cómo *big data* puede ser una herramienta clave para la construcción de contrafácticos.

Big data no necesariamente es mucha información

No solo *big data* no es todos los datos sino que a veces ni siquiera es mucha información. A modo de ejemplo, la pobreza en el Gran Buenos Aires se mide con una encuesta periódica oficial (la Encuesta Permanente de Hogares) de aproximadamente 3.500 hogares, que representan a los aproximadamente 4 millones de hogares de la región. Es la estructura científica de la misma la que garantiza que unos pocos hogares pueden cumplir adecuadamente la tarea de representar una población mucho mayor. Suena tentador pensar que una encuesta que responden millones de personas en las redes sociales es mucho mejor que una encuesta tradicional. Como adelantásemos, la naturaleza anárquica y espontánea de *big data* hace que sus datos sean incomparables con los de una encuesta o experimento científicamente diseñados.

Detrás de la idea de que "muchos datos es mejor que pocos", está la Ley de Grandes Números (LGN), uno de los pilares de la estadística convencional. Sin entrar en tecnicismos, la LGN es lo que garantiza que las chances de que lanzar una moneda salga cara puedan aprenderse de tirar repetidamente una moneda y contar la proporción de veces que sale cara. Implícito en este razonamiento hay dos requisitos de la LGN: 1) que siempre se tire la misma moneda, 2) que en el transcurso de tirar la moneda no se aprenda a hacerlo. Técnicamente, el primer requisito es el de idéntica distribución y el segundo el de independencia. Es decir, para que la LGN permita aprender probabilidades de repetir un experimento infinitas veces es necesario que estos experimentos sean siempre los mismos e independientes entre sí.

> No solo *big data* no es todos los datos sino que a veces ni siquiera es mucha información (...) la naturaleza anárquica (...) de *big data* hace que sus datos sean incomparables con los de una encuesta o experimento científicamente diseñados.

El paradigma de muestreo al azar intenta garantizar que dichas condiciones se cumplan en una muestra así diseñada o, como ocurre en la práctica, que un muestreo complejo cumpla con las condiciones ideales que más lo asemejan a este paradigma ideal.

Por el contrario, *big data* es casi la negación del ideal de "independencia y distribución idéntica". Es su naturaleza espontánea lo que hace que sus datos sean fuertemente dependientes y que refieran a poblaciones marcadamente heterogéneas.

En un sorpresivo paper reciente, Xiao Li Meng (2016), director del departamento estadística de la Universidad de Harvard, muestra cómo, por las cuestiones discutidas, los millones de datos de *big data* tienen serias dificultades en predecir eventos como la elección del presidente Trump, fenómeno al que refiere como "la paradoja de *big data*": muchos datos no necesariamente se traducen en mucha información.

El desafío de la ética, la transparencia y la privacidad

El manejo de la cuestión pública es necesariamente multidimensional y con objetivos múltiples, muchos de ellos contradictorios. Así, muchas acciones justificables y deseables en pos de un objetivo son descartadas por atentar contra otros.

Un ejemplo concreto lo constituye el caso de la política del gobierno noruego, que en pos de la transparencia exigió históricamente que los ingresos de todos sus ciudadanos fuesen públicos. Hasta no mucho tiempo atrás, cualquier ciudadano podía consultar el ingreso de cualquier otro, a través de un simple trámite burocrático.

Recientemente, el avance de la tecnología permitió que el ingreso de todos los noruegos fuese accesible a través de una plataforma online, y no mucho después, existió una app que permitía cosas tales como, con un simple click, conocer el ingreso de todos los contactos de Facebook de cualquier persona. En muy pocos días se desató un escándalo masivo de episodios de *bullying* (de los más ricos a los más pobres) y un vendaval de peleas por comparaciones de ingresos en personas trabajando en instituciones similares. A la luz de estos conflictos, el gobierno noruego dio marcha atrás con esta política, requiriendo, una vez más, volver a algunas trabas burocráticas para que se pudiese tener acceso al ingreso de un ciudadano.

Este simple ejemplo muestra que las acciones deseables en pos de un objetivo honesto, como la transparencia, entran en conflicto con otro, de la misma estatura moral, como la privacidad. El episodio reciente de Cambridge Analitica puede ser también entendido en este marco: en pos de la eficiencia es tal vez deseable que se difunda cualquier tipo de información, pero atentando contra un valor relevante como la privacidad y la confiabilidad de las instituciones.

Los avances de *big data* y *machine learning* provienen fundamentalmente desde el sector privado, en donde en pos de la eficiencia y los beneficios hay espacio para experimentar con una variedad de estrategias. El sector público, por su conformación, debe velar por otros objetivos adicionales, como la transparencia, la fe pública o los principios éticos más elementales y debe ser naturalmente cauto en la adopción irresponsable de tecnologías que atenten contra ellos. Velar por estos principios y crear las instituciones para regular su operatoria, es un desafío mayor que enfrenta la adopción de *big data* en el sector público, desafío de idéntica complejidad y relevancia que el planteado por lo más profundo de la tecnología.

El desafío del consenso y el de la comunicabilidad

"Las líneas de pobreza son construcciones tan políticas cómo científicas" dijo Angus Deaton, premio Nobel en economía 2015, ilustrando la delicada conexión que hay entre la estadística social y la práctica de la política. Más concretamente, no existe ninguna forma indiscutible de medir la pobreza, como consecuencia de que no existe una forma unívoca de definir qué significa ser pobre. De modo que, en la práctica, cualquier medición de la pobreza (desde el enfoque de líneas hasta los recientes avances en pobreza multidimensional) son el resultado de una delicadísima negociación entre cuestiones económicas, sociales, culturales, computacionales, informativas y políticas. Nadie argumenta que la medición de pobreza en base a ingresos es buena, ni siquiera correcta, tan solo que es útil, es decir que satisfacen algún fin pragmático que surge de negociar las considerables discrepancias conceptuales, las dificultades técnicas de relevar información y la necesidad de que esta medición resulte comparable en el tiempo y en el espacio. En este contexto, en lo que refiere a la medición de la pobreza (por discutir un ejemplo relevante) el rol del Estado no se remite a la implementación de un ejercicio estadístico-computacional sino a dotar a las cifras sociales de una credibilidad producto de un consenso técnico, operativo y conceptual, que garantice la relevancia y la transparencia de la estadística pública.

Más aún, es casi un hecho estilizado de la estadística aplicada que los métodos usados en la práctica no necesariamente son aquellos con mejores

propiedades teóricas sino aquellos que son percibidos como transparentes y comunicables por toda la comunidad relevante, que incluye a los técnicos, pero también a los políticos, periodistas y comunicadores en general, cuando no a toda la población. A modo de ejemplo, la "tasa de incidencia" de la pobreza (proporción de personas u hogares con ingresos inferiores a una línea de pobreza) permite construir otras medidas de pobreza tal vez mejores (como la tasa de profundidad de pobreza), pero que exigen un mayor esfuerzo comunicacional, que las relega a un segundo plano, cuando desde un punto de vista estrictamente técnico sean preferibles.

> Lamentablemente, la conjunción de *big data* y *machine learning* tiende a favorecer la conformación de auténticas "cajas negras", que pueden resultar relevantes en muchos ámbitos (particularmente, en el sector privado) pero inaceptable política y comunicacionalmente en el sector público.

Lamentablemente, la conjunción de big data y machine learning tiende a favorecer (a veces imprudentemente), la conformación de auténticas "cajas negras", que pueden resultar relevantes en muchos ámbitos (particularmente, en el sector privado) pero inaceptable política y comunicacionalmente en el sector público.

Las discusiones que tuvieron lugar en relación a la confiabilidad de la medición de la pobreza en Argentina en los últimos años proveen un claro ejemplo, en donde una medida algorítmicamente trivial (como la usada en el enfoque de línea de pobreza) fue puesta en consideración por dudar de los datos usados para su cómputo. Es cuestión de imaginar el peligroso "cocktail" político y comunicacional que resultaría de calcular la pobreza con datos sospechosos y oscuros algoritmos de *machine learning*.

La adopción de estrategias de *machine learning* y *big data* en el sector público demanda un enorme desafío, que implica sostener una comunidad científica dispuesta a interactuar activamente con la complejidad de la gestión pública y que, en conjunto y en forma multi e interdisciplinar, permitan dotar a los algoritmos de la credibilidad y comunicabilidad que exige la práctica diaria de la gestión pública.

No se trata de favorecer innecesariamente métodos simples sino de lo contrario: de dotar a la complejidad de una credibilidad social, que permita al sector público aprovecharse de las más recientes tecnologías, pero a la vez dejando tranquilos a los ciudadanos que solo van a aparecer las ventajas de hacerlo y ningún de sus problemas más delicados.

Comentarios finales

El tiempo dirá si el fenómeno de *big data* es un auténtico cambio de paradigma, un avance técnico o una moda pasajera. Esta nota intenta aportar una visión en donde, como con cualquier tecnología disruptiva, hay ventajas, limitaciones y, fundamentalmente, zonas grises. Los casos discutidos hablan del enorme potencial de *big data* en la cosa pública. Los desafíos discutidos sugieren que la adopción ciega de datos masivos y algoritmos puede conducir a crear falsas expectativas en los usuarios, cuando no a serios errores que se manifiestan a través de sesgos y omisiones.

> Tal vez *new data* sea una mejor forma de referir al fenómeno de datos recientes. Es decir, esta terminología resalta que, más allá de la masividad, lo que verdaderamente tiene *big data* para aportar sea la posibilidad de observar información hasta hace muy poco tiempo impensable, acerca de la forma en la que operan las sociedades, en términos de sus preferencias, acciones o mecanismos.

Tal vez *new data* sea una mejor forma de referir al fenómeno de datos recientes. Es decir, esta terminología resalta que, más allá de la masividad, lo que verdaderamente tiene *big data* para aportar sea

la posibilidad de observar información hasta hace muy poco tiempo impensable, acerca de la forma en la que operan las sociedades, en términos de sus preferencias, acciones o mecanismos. En todo caso, big data no es "más de lo mismo". Tal vez el principal error consiste en pensar a los nuevos datos como si fuesen los viejos, pero en una escala mayor.

La posibilidad de que *big data* ofrezca sus principales beneficios y pocas de sus limitaciones dependerá de un abordaje que va mucho más allá de lo meramente técnico y algorítmico. Lidiar con sus desafíos y limitaciones requiere un enfoque abarcativo, multidisciplinar, que sopese las ventajas técnicas con las limitaciones conceptuales, éticas y operativas del fenómeno.

Referencias

- Blumenstock, J., Cadamuro, G., On, R., 2015, Predicting poverty and wealth from mobile phone metadata, Science, 27, Vol. 350, Issue 6264, pp. 1073-1076
- Cavallo, Alberto, and Roberto Rigobon. 2016. The Billion Prices Project: Using Online Prices for Measurement and Research, Journal of Economic Perspectives, 30 (2): 151-78.
- Einav, Liran, Dan Knoepfle, Jonathan Levin, and Neel Sundaresan. 2014. Sales Taxes and Internet Commerce. American Economic Review, 104 (1): 1-26.
- Gasparini, L., Cicowiez, M. y Sosa Escudero, W., 2013, Pobreza y desigualdad en America Latina: conceptos, herramientas y aplicaciones, TEMAS Grupo Editorial, Buenos Aires.
- Mayer-Schönberger, V y Cukier, K., 2013, Big data. La revolución de los datos masivos. Editorial Turner, Madrid.
- Meng, X. L, 2016, Statistical paradises and paradoxes in big data (I): Law of large populations, big data paradox, and the 2016 US presidential election, The Annals of Applied Statatics, Volume 12, Number 2 (2018), 685-726.
- O'Neil, C., 2018, Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y amenaza la democracia, Capitán Swing, Madrid.
- Sosa Escudero, W. 2019, Big data, 5ta edición, Siglo XXI Editores, Buenos Aires.
- Sosa Escudero, W. 2020, Big data: Desafíos para la política pública, mimeo, CLAD.

Contacto

Organización Internacional del Trabajo
Oficina de la OIT para el Cono Sur de América Latina.
Santiago de Chile

T: (56-2) 2580-5500
E: santiago@ilo.org
W: ilo.org/santiago