

***Big data*: desafíos para la política pública**

Walter Sosa Escudero

Walter Sosa Escudero

Doctor en Economía (University of Illinois en Urbana-Champaign) y Licenciado en Economía (Universidad de Buenos Aires -UBA). Se especializa en econometría y estadística aplicada a cuestiones sociales. Sus trabajos de investigación se encuentran en publicaciones nacionales e internacionales, incluyendo Journal of Econometrics, Journal of Economic Inequality y Econometric Theory. Es profesor plenario y director del Departamento de Economía de la Universidad de San Andrés. Agrega a la investigación y docencia una agenda activa de divulgación científica sobre el rol de los datos y las estadísticas en la sociedad. Fue orador de TEDxRíodelaPlata 2013 y es autor de "*Big Data*: breve manual para conocer la ciencia de datos que ya invadió nuestras vidas" y "Qué es (y qué no es) la estadística: usos y abusos de una disciplina clave en la vida de los países y las personas", además de escribir frecuentemente para los principales diarios argentinos. Ha recibido los premios Fulvio Pagani/ARCOR, Bernardo Houssay, Konex, el Premio EGSO como el mejor profesor del Doctorado en Economía de la Universidad de Illinois y fue elegido por la Fundación BGH como uno de los 100 argentinos innovadores. Fue presidente de la Asociación Argentina de Economía (2012-2014). Es investigador principal del Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) y miembro titular de la Academia Nacional de Ciencias Económicas.

Las comunicaciones con el autor pueden dirigirse a:
E-mail: wsosa@udesa.edu.ar

Big data: desafíos para la política pública*

Este artículo revisa el fenómeno de big data y su potencial para el uso en la cuestión pública. Argumenta que big data no es más de lo mismo y que plantea exigencias que, si son entendidas como desafíos a sortear, pueden abrir la puerta a las enormes ventajas de los datos masivos y los algoritmos, garantizándole al ciudadano que no aparecerán sus costados menos deseables. Se adopta una postura científica y se remarca el enorme potencial del fenómeno de los datos masivos, sus principales deficiencias y, fundamentalmente, el terreno intermedio conformado por desafíos e incertidumbres propias de un fenómeno dinámico y todavía inestable.

Palabras clave: Informática; Tecnología de la Información; Tecnología de las Comunicaciones; Política Pública; Sector Público; Perspectivas

Big Data: Challenges por Public Policy

This article explores the phenomenon of big data and its potential for use in the public sector. It argues that big data is not just “more of the same”, but that it face challenges that can open the door to the enormous advantages in the use of algorithms and massive data, guaranteeing to the citizens that its less desirable effects will not appear. A scientific perspective is adopted to emphasize the enormous potential of big data, its main limitations, and, most importantly, the intermediate arena formed by challenges and uncertainties typical of a dynamic and still unstable phenomenon.

Key words: Informatics; Information Technology; Communication Engineering; Public Policy; Public Sector; Perspectives

Introducción

Big data es un concepto reciente que refiere al fenómeno de datos masivos producto de la interacción con dispositivos interconectados, como teléfonos celulares, dispositivos de GPS, sensores, tarjetas de crédito, entre otros. La contracara de los datos son los algoritmos o métodos computacionales, matemáticos y estadísticos utilizados para sistematizar y analizarlos. Aprendizaje automático, inteligencia artificial y estadística clásica son las disciplinas que conforman este

Recibido: 14-11-2019 y 09-03-2020 (segunda versión). Aceptado: 10-03-2020.

(*) Versión revisada de la Conferencia Magistral presentada en el XXIV Congreso Internacional del CLAD sobre la Reforma del Estado y de la Administración Pública, Buenos Aires, Argentina, 12 al 15 de noviembre de 2019, con el título original “Big data: casos, perspectivas y desafíos para la gestión pública”. Se agradecen los comentarios e ideas de Lucila Berniell.

conglomerado analítico que es tan o más importante que los datos. Asimismo, disciplinas aparentemente lejanas, como la comunicación visual, el diseño y la comunicación, también cumplen un rol crucial en esta nueva visión interactiva que plantea el fenómeno de los datos masivos. Finalmente, el conocimiento específico de todas las disciplinas que convocan al estudio sistemático de datos -desde las más duras como la física, hasta las humanísticas como la lingüística- cumplen también un papel fundamental. El concepto “ciencia de datos” es tanto una forma ligera y redundante de referir al análisis de datos, como un indicio de que es relevante una nueva denominación para un fenómeno esencialmente distinto, que requiere una auténtica visión panóptica y altamente interactiva de procesos otrora desintegrados.

Como es el caso de cualquier tecnología disruptiva, las posturas acerca del fenómeno de *big data* son polares. Por un lado, están los que opinan que se trata de un auténtico cambio de paradigma que tornará rápidamente obsoletas a muchas prácticas atávicas, como las encuestas, los experimentos y, también, el propio método científico. Por el otro, están los que ven en *big data* otra moda pasajera y tienden a adoptar posturas conspirativas, resaltando los difíciles problemas de transparencia y privacidad que la disponibilidad de datos conlleva. El texto de Mayer-Schönberger y Cukier (2013) es un claro representante de la visión más optimista y el de O’Neil (2018), de la visión más escéptica.

Este artículo revisa el fenómeno de *big data* y su potencial para el uso de la cuestión pública. Pretende tener una postura científica y señala su enorme potencial, sus principales deficiencias y, fundamentalmente, el terreno intermedio conformado por desafíos e incertidumbres propias de un fenómeno dinámico y todavía inestable. Una parte de las visiones ofrecidas en este artículo se basa en el reciente libro del autor (Sosa Escudero, 2019).

¿De qué hablamos cuando hablamos de *big data*?

Cualquier definición precisa de un fenómeno todavía cambiante parece condenada a una prematura obsolescencia. La mera etimología de la frase “*big data*” remite a datos masivos. Pero es importante remarcar que la masividad es, tal vez, la menos relevante de las características definitorias de este fenómeno. Los datos de *big data* surgen de la interacción espontánea con tecnologías interconectadas, como teléfonos celulares, dispositivos de GPS o redes

El fenómeno de *big data* tiene que ver tanto con la masividad de datos no estructurados, espontáneamente generados por interactuar con dispositivos interconectados, como con los algoritmos y métodos destinados a estudiarlos.

sociales. Su naturaleza espontánea contrasta con la estructura subyacente en fuentes de datos tradicionales, como una encuesta o experimento científicamente diseñado.

Consecuentemente, los datos de *big data* no son “más de lo mismo”, sino un fenómeno completamente distinto. La espontaneidad y consecuente falta de estructura explican tanto la masividad del fenómeno como sus principales limitaciones. A modo de ejemplo, el diseño científico de la Encuesta Permanente de Hogares para el Gran Buenos Aires, que se implementa en unos 3.000 hogares, permite extrapolar resultados a los aproximadamente 4 millones de hogares. Estos datos son incomparables a los miles de respondientes a una encuesta *online* implementada en las redes sociales, toda vez que existe alguna razón por la que las personas: a) deciden usar una red social, b) deciden seguir a un usuario en particular y c) deciden responder a una encuesta. En definitiva, los datos de *big data* no son una ampliación trivial de los datos de fuentes tradicionales y, consecuentemente, plantean nuevos desafíos que consisten, mayoritariamente, en dotar de estructura a datos anárquicos y espontáneos.

Tan importante como los datos son los métodos numéricos, matemáticos, estadísticos y computacionales usados para ordenar, sistematizar y analizar los datos. El “Machine learning”, la inteligencia artificial y, también, la estadística clásica cumplen un rol crucial en la revolución de datos.

En síntesis, el fenómeno de *big data* tiene que ver tanto con la masividad de datos no estructurados, espontáneamente generados por interactuar con dispositivos interconectados, como con los algoritmos y métodos destinados a estudiarlos.

Dos casos

Pobreza en Ruanda

Ruanda es un país extremo en lo que se refiere al flagelo de la pobreza. No es relevante argumentar demasiado acerca de la intensidad del fenómeno ni de la relevancia de diseñar políticas destinadas a mejorar el bienestar de su población. La medición moderna del bienestar implica llevar a cabo un complejo sistema de encuestas a fin de garantizar que las mediciones obtenidas sean comparables para varios períodos y regiones. A modo de ejemplo, la medición basada en el “enfoque de líneas” (Gasparini, Cicowiez y Sosa Escudero, 2013) requiere una encuesta regular de ingresos y otra de

Mientras que el US Bureau of Labor Statistics de Estados Unidos releva unos 80.000 precios en un período de uno a dos meses, los métodos de *scrapping* permiten 5 millones de precios de 300 negocios en 50 países en tan solo un día.

precios y consumos que permitan observar si los ingresos de un hogar superan el valor de una “línea” debajo de la cual se considera que un hogar es pobre.

El tipo de esfuerzo institucional que requiere un sistema organizado de encuestas regulares está muy lejos de las posibilidades de un país como Ruanda. Blumenstock, Cadamuro y On (2015) construyeron un modelo algorítmico que permite predecir la intensidad de la pobreza en Ruanda con base en el uso de teléfonos celulares. El método se basa en una pequeña encuesta de bienestar que es utilizada para entrenar un algoritmo de *machine learning* que permite predecir el bienestar (y la pobreza) con base en cuán intensa y frecuentemente se usan los celulares. Si bien perfectible y todavía conjetural, este tipo de esfuerzos permitiría obtener una medición de la pobreza para varias regiones de ese país, y comparable en el tiempo, lo que constituiría una importante herramienta para el diseño y monitoreo de políticas sociales.

Índices de precios

La existencia de numerosos canales de venta *online* provee una oportunidad única para acceder inmediatamente a una enorme variedad de precios, con una amplia cobertura geográfica y en tiempo prácticamente virtual. Históricamente la naturaleza idiosincrática de estos canales de venta planteaba un desafío a la hora de extraer, validar y sistematizar esta información. Más concretamente, los sitios web de un supermercado están diseñados para conveniencia del consumidor y no del analista. Los precios no se ofrecen en listas sistemáticas sino obedeciendo a cuestiones estéticas y funcionales que facilitan la compra *online*. Afortunadamente, como señalan Cavallo y Rigobon (2016), las tecnologías de *web scrapping* han avanzado lo suficientemente rápido como para que esto no sea un inconveniente. Este tipo de “robot computacional” accede a los códigos detrás de las páginas web y, con un mínimo de monitoreo y aprendizaje, puede construir bases de datos de precios a partir de información no sistemática, como la de un sitio de ventas *online*.

El así llamado Proyecto del Billón de Precios, con base en MIT, recoge precios de una enorme variedad de productos en varios países. A fines de resaltar las ventajas operativas de estas estrategias los autores reportan, que mientras que el US Bureau of Labor Statistics de Estados Unidos releva unos 80.000 precios en un período de uno a dos meses, los métodos de *scrapping* permiten 5 millones de precios de 300 negocios en 50 países en tan solo un día.

Cuatro desafíos

Los casos anteriores son solo dos ejemplos concretos de uso de datos espontáneamente producidos (de intensidad de uso de celulares, en el primer caso, y de precios de tiendas *online*, en el segundo) que se contraponen en su esencia con los datos provenientes de fuentes tradicionales, como una encuesta de hogares o un experimento científico. Es la naturaleza espontánea, y muchas veces anárquica, de los datos de *big data* lo que constituye su principal ventaja y a la vez su principal fuente de limitaciones. Esta sección discute algunos desafíos concretos que plantea la naturaleza espontánea de *big data* a su uso en la cuestión pública.

Big data no es (ni será) todo los datos

Alguna práctica sobre-entusiasta de *big data* tiende a pensar que el fenómeno de datos masivos implica que se está muy cerca de llegar a una situación en donde se dispondrá de “todos los datos”. Es decir, no tiene sentido en insistir en la vieja práctica de las encuestas que intentan aproximar una población a través de una parte cuando, *big data* mediante, es posible acceder directamente a la población. Esta visión, errada, sugiere que el rol clásico de la estadística tiene los días contados a la luz de que ahora es posible atacar directamente a la población, sin intermediar ninguna muestra ni diseño muestral.

El flamante premio Nobel a Esther Duflo, Abhijit Banerjee y Michael Kremer 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 le administró dicha droga con *exactamente* la misma persona en la circunstancia en la que no se le 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. Por ello, se observan 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,

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.

de modo que, a fines de los objetivos del experimento, lo único que 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 contrafácticos. A modo de ejemplo, en plena época de *big data*, *machine learning* e inteligencia artificial, la evaluación de una política como la Asignación Universal por Hijo de la República Argentina requiere de delicados mecanismos estadísticos, porque la misma no fue asignada al azar. En este caso, la comparación directa de la *performance* personas que recibieron esta política y las que no, es como comparar peras con manzanas, toda vez que esas personas difieren en muchas más cosas relevantes que el hecho de haber sido receptoras de una política. Es decir, *big data* no puede ser todos los datos. En modo optimista tal vez sea capaz de revelar, a lo sumo, la mitad de los datos producto de las acciones, pero no de sus contrafácticos por definición. Es crucial remarcar que 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) sobre los efectos causales del impuesto a las ventas, es una gran ilustración de cómo usar datos de *big data* 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 puedan cumplir adecuadamente la tarea de

representar una población mucho mayor. Por otro lado, cualquier *celebrity* menor cuenta sus seguidores en Twitter o Instagram de a cientos de miles, cuando no millones. Suena tentador pensar que una encuesta que responden millones de personas es mucho mejor que una encuesta tradicional, implementada con formularios y encuestadores que recogen unos pocos cientos o miles de datos. Como se adelantó, 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 y que 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 y 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. 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 que más lo asemejan a ese 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, Xiao Li Meng (2018), director del departamento estadístico de la Universidad de Harvard, muestra cómo, por las cuestiones discutidas, los millones de datos de *big data* tiene 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.

Nuevamente, no se trata de concluir que los datos de *big data* son inútiles, sino que no pueden ser inocentemente tomados como si fuesen una versión extendida de los datos tradicionales. Proveer a estos datos de alguna estructura que permita aprovechar su potencial,

La masividad de *big data* puede ser un útil complemento de las fuentes tradicionales como las encuestas o la contabilidad nacional, que por su planificación metódica constituyen una importantísima fuente confiable de información para la toma de decisiones.

es un crucial desafío para la política pública. La masividad de *big data* puede ser un útil complemento de las fuentes tradicionales como las encuestas o la contabilidad nacional, que por su planificación metódica constituyen una importantísima fuente confiable de información para la toma de decisiones.

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 conocer el ingreso de todos los contactos de Facebook de cualquier persona con un simple clic. 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 de Cambridge Analítica también puede ser 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. Por su conformación, el sector público

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.

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 como 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) es el resultado de una delicadísima negociación entre cuestiones económicas, sociales, culturales, computacionales, informativas y políticas. Nadie argumenta que la medición de pobreza con base en ingresos es buena, ni siquiera correcta, tan solo que es *útil*. Es decir, satisface 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

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, pero que son inaceptables política y comunicacionalmente en el sector público.

(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, lo que las relega a un segundo plano aun cuando desde un punto de vista estrictamente técnico sean preferibles.

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 que son inaceptables política y comunicacionalmente en el sector público. La álgida discusión reciente sobre la medición de la pobreza en Argentina provee un claro ejemplo. Es cuestión de imaginar el peligroso “cocktail” político y comunicacional que resultaría de alimentar oscuros algoritmos con datos sospechosos para medir la pobreza, si la misma fuese medida usando algoritmos de *machine learning*.

Que la eficiencia de *machine learning* y *big data* pueda implementarse en el sector público requiere de 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 interdisciplinaria, 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 en que solo van a aparecer las ventajas de hacerlo y no sus problemas más delicados.

Comentarios finales

No existe ámbito ajeno a la revolución de los datos y, claramente, el sector público no es una excepción. Por el contrario, *big data* ofrece una oportunidad única para mejorar sustancialmente la gestión pública. Este artículo argumenta que *big data* no es más de lo mismo y que plantea exigencias que, si son entendidas como desafíos a sortear, pueden abrir la puerta a las enormes ventajas de los datos masivos y los algoritmos, garantizándole al ciudadano que no aparecerán sus costados menos deseables.

Las cuestiones de transparencia, comunicabilidad y ética plantean dificultades cuya complejidad es comparable a las propias de las cuestiones más duras detrás de la tecnología algorítmica. La adopción provechosa de *big data* en el sector público requiere una comunidad madura, que permita que cada sector entienda honestamente la complejidad del otro. El sector público no puede adoptar pasivamente cualquier tecnología de moda, ni el sector tecnológico puede pretender ignorar la complejidad del entramado social. Si estas complejidades son explicitadas, negociadas y debidamente canalizadas, el potencial de *big data* en la cosa pública es enorme.

Bibliografía

- Blumenstock, Joshua; Cadamuro, Gabriel; y On, Robert (2015), "Predicting Poverty and Wealth from Mobile Phone Metadata", en *Science*, Washington, Vol. 350 N° 27, Issue 6264, pp. 1073-1076.
- Cavallero, Alberto y Rigobon, Roberto (2016), "The Billion Prices Project: Using Online Prices for Measurement and Research", en *Journal of Economic Perspectives*, Vol. 30 N° 2, pp. 151-178.
- Deaton, Angus (2015), *El gran escape: salud, riqueza y los orígenes de la desigualdad*, México, Fondo de Cultura Económica.
- Einav, Liran; Knoepfle, Dan; Levin, Jonathan; y Sundaresan, Neel (2014), "Sales Taxes and Internet Commerce", en *American Economic Review*, Vol. 104 N° 1, pp. 1-26.
- Gasparini, L.; Cicowiez, M.; y Sosa Escudero, W. (2013), *Pobreza y desigualdad en América Latina: conceptos, herramientas y aplicaciones*, Buenos Aires, Temas Grupo Editorial; Universidad Nacional de La Plata. Facultad de Ciencias Económicas. Centro de Estudios Distributivos, Laborales y Sociales.
- Mayer-Schönberger, Viktor y Cukier, Kenneth (2013), *Big data. La revolución de los datos masivos*, Madrid, Editorial Turner.
- Meng, Xiao-Li (2018), "Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election", en *The Annals of Applied Statistics*, Vol. 12 N° 2, pp. 685-726.

O'Neil, Cathy (2018), *Armas de destrucción matemática. Cómo el big data aumenta la desigualdad y amenaza la democracia,*

Madrid, Capitán Swing.
Sosa Escudero, Walter (2019), *Big data*, Buenos Aires, Siglo XXI Editores. 5. ed.