

REFLEXIONES SOBRE LOS NUEVOS DESAFÍOS ÉTICOS QUE PLANTEA EL USO DE GRANDES BASES DE DATOS (BIG-DATA) EN INVESTIGACIÓN¹

María Teresa Blaconá*

Resumen. Desde hace unos años, las grandes bases de datos se han convertido en una herramienta fundamental para investigaciones en varias áreas de la ciencia, entre las que se pueden distinguir: la biología y salud, las ciencias sociales (a partir de la información personal en redes sociales), negocios y finanzas. Esta irrupción ha venido acompañada de nuevas o adaptadas metodologías en disciplinas como la estadística y los sistemas de información entre otras. Esto ha llevado a que se revisen los conceptos éticos tradicionales de confidencialidad, privacidad y anonimato debido a las nuevas dimensiones que adquieren al manejar grandes bases de datos. En este trabajo se discuten algunos aportes realizados en los últimos años sobre el tema con la recomendación de que éstas se tengan en cuenta cuando se utilizan en investigación. Ante tantos avances tecnológicos y metodológicos es importante para la comunidad científica que se posibilite su discusión en foros estadísticos tanto sea desde un punto de vista académico, empírico y/o gubernamental. Además es imprescindible presentar esta problemática a estudiantes universitarios junto con una divulgación accesible y clara para la población general.

Palabras Clave: Anonimato; Privacidad; Confidencialidad.

¹ Agradecimientos a la Dra. Liliana A. Picardi, la Mag. Cristina Cuesta y el Mag. Javier Bussi por sus comentarios y sugerencias.

REFLECTIONS ON THE NEW ETHICAL CHALLENGES PROVIDED BY THE USE OF BIG DATA BASES (BIG-DATA) IN RESEARCH

Abstract. For many years large data base have become a fundamental tool for investigation in various scientific fields, including: Biology and Health, the Social Sciences (through personal information on social networks), Bussiness and Finance. This irruption has been accompanied by new o adapted methodologies in disciplines like Statistics and Information Technology, among others. This has led to the revisión of the traditional ethical concepts of Confidentiality, Privacy and Anonymity due to the new dimensions that are acquired when handling large data bases. Contributions made in recent years on the subject in question are discussed in this paper with the recommendation that they should be considered when used in research. Faced with so many technological and methodological advances it is important for the scientific coomunity to enable the discussion of the implications of large data bases in statistical fóruns both from an academic, empirical and/or governmental point of view. It is also essential to present this problem to university students together with an accesible and clear disclosure for the general popilation.

Keywords: Privacy; Anonymity; Confidentiality

Original recibido el 01/11/2019.

1. Introducción

Desde hace unos años, las grandes bases de datos con información personal se han convertido en una herramienta fundamental para investigaciones en varias áreas de la ciencia, entre las que se pueden distinguir: la biología y salud, las ciencias sociales (a partir de la información personal en redes sociales), los negocios y las finanzas. Esta irrupción ha venido acompañada de nuevas o adaptadas metodologías en disciplinas como la estadística y los sistemas de información.

Los avances metodológicos de gran impacto necesitan ser mirados críticamente, comprendidos en su alcance y examinados desde la utilidad que aportan. Se debe dar la posibilidad de interrogarse sobre esta utilidad y, sobre todo, de sus limitaciones éticas.

Big-Data se ha vuelto tan grande e importante porque en este mundo globalizado cada vez hay una mayor demanda de información sobre consumidores y votantes (Nichols, 2018). Gestionar los datos en forma responsable ya no es opcional. Los datos hoy son un activo de un valor incalculable para los profesionales y se han convertido en una materia prima valiosa, a la altura del petróleo o el oro. Es por esto que la privacidad del usuario debe ser una prioridad (y una línea roja a no traspasar) para todas las empresas que buscan beneficiarse con ello (Vogelberg, 2019). Es en este sentido que Hirsch y King (2015) alertan sobre que *Big-Data* es el nuevo “petróleo” del cual nuestra sociedad depende. Mientras el petróleo produce dramáticos derrames, *smog* y contribuye a acelerar los cambios climáticos, *Big-Data* también puede conducir a derrames, violación de privacidad, polución de identidad y daños por discriminación. Cuando han sido usados a gran escala, indiscriminadamente, han producido no sólo grandes beneficios, sino también significativas (a veces involuntarias) tensiones en aquellas sociedades en las que se puede entrar en conflicto con las regulaciones vigentes.

Vayena, Grasser, Wood, O'Brian y Altman (2015) reconocen el inmenso valor que tienen las investigaciones en *Big-Data* y creen que las nuevas fuentes de datos en gran escala deberían estar disponibles de tal manera que todo su potencial pueda ser realmente aprovechado. Dado los sustanciales beneficios de las investigaciones con *Big-Data*, la pregunta central que se debería hacer hacia el futuro no es si las grandes bases de datos deberían estar disponibles, sino cómo se puede capturar su valor de tal forma que se respeten los principios fundamentales de la ética y la privacidad. Estos autores argumentan que las investigaciones y las conclusiones que son obtenidas con *Big-Data* presentan nuevos riesgos ya que las actuales normas regulatorias son inadecuadas para abordarlos. Recomiendan adaptar estas normas para que permitan coleccionar, usar y compartir *Big-Data* siguiendo principios éticos, normas comunitarias en investigación y expectativas del sujeto humano. La principal recomendación sería reconocer los derechos humanos de los participantes en la producción del conocimiento científico. Los derechos humanos en la ciencia se refieren a las obligaciones que los gobiernos y otros actores deben ejercer para proteger y comprometer su participación en todas las etapas que constituyen el proceso de la investigación científica.

La metodología de investigación utilizando *Big-Data*, tendiente a la identificación

de patrones de comportamiento a partir de grandes volúmenes de datos, presenta características especiales que habrá que tener en cuenta. La naturaleza de algunos estudios de *Big-Data* está libre de hipótesis, lo que hace muy difícil aplicar mecanismos epistemológicos convencionales (García León, 2018).

Las investigaciones con *Big-Data* son algo nuevo, ya que generalmente coleccionan datos en gran escala que frecuentemente combinan aspectos de investigación, experimentación y entretenimiento. Hoy en día las investigaciones con *Big-Data* son, por su naturaleza, exploratorias. Algunas de estas investigaciones exploratorias tienen un propósito, mientras que otros proyectos se parecen a “expediciones de pesca”, prácticas que son consideradas no éticas en muchas disciplinas. El hecho destacable es que se puede coleccionar un gran número de información, analizarla y volver a analizarla hasta que un resultado se vuelva aparente según preconcepciones previas (Garfinkel, 2015).

A medida que la tecnología avanza, los desafíos éticos de *Big-Data* crecen y surgen preguntas que la sociedad nunca se había hecho antes. No es ningún secreto que los datos de pacientes guardados electrónicamente condujeron a que se presenten en los últimos años un gran número de nuevos problemas. Ataques cibernéticos llevaron a violaciones de datos que han comprometido la privacidad de pacientes de EEUU. Solamente en el año 2017, se identificaron 477 violaciones a organizaciones de cuidados de la salud que afectó 5,6 millones de registros de pacientes (Ayers, 2018).

También es un hecho conocido que con algoritmos y conocimientos de computación es posible describir y determinar perfiles personales. Esto conlleva a incrementar temas éticos y eventualmente legales, especialmente con la decisión automática de las máquinas basadas en modelos analíticos y algoritmos. Estos mecanismos podrían desencadenar acciones en tiempo real y/o posteriores que afectarían la justicia social y económica de una persona o una comunidad entera (Oliveira, 2018).

En estos días, las investigaciones tanto en instituciones académicas como en el sector privado en negocios, no están sujetas a una Regla Común (como la adoptada por el Institutional Review Board-IBR de EEUU en 1979 y sus actualizaciones), lo cual condujo a un reglamento estándar en ética en investigación y llevó a un entendimiento básico de los temas entre investigadores que realizaban investigaciones con sujetos humanos; pero no ha sido perfecto, ha fallado efectivamente en balancear el progreso de la investigación y la protección de las personas (Clifton, 2015). El entendimiento básico manejando una amplia fuente de datos, desde datos comerciales masivos, base de datos gubernamentales, datos personales de redes sociales disponibles *on line*, no brinda ninguna oportunidad de involucrar directamente a las personas participantes, obtener su consentimiento y además informarles de las actividades de la investigación en la cual se los está involucrando.

El desafío es ajustar la delimitación de los datos focalizados en la investigación al problema en estudio, porque existen características éticas y legales que determinarán si la sociedad se puede beneficiar de las tremendas oportunidades, escondidas en los datos exhaustivos de gobiernos, instituciones al cuidado de la salud, escuelas, redes sociales, mientras al mismo tiempo se proteja la privacidad, ecuanimidad, igualdad

de las personas y la integridad del proceso científico (Tene y Polonitsk, 2015). A esto, Croll (2012) lo llamó “el tema más grande de derechos humanos de nuestro tiempo”, porque los seres humanos se vuelven “transparentes” a la luz de la información obtenida sin su consentimiento. Esta “transparencia” forzada lleva a replantarse los conceptos éticos tradicionales de anonimato, privacidad y confidencialidad, entre otras cosas, por la nueva dimensión que adquieren al manejar grandes bases de datos.

En este trabajo se han resumido estos conceptos en distintas secciones. En la segunda sección se enuncian los conceptos básicos para el tratamiento ético de los datos. En la tercera sección se presentan tres grandes áreas donde con más frecuencia se utilizan grandes bases de datos, como lo son: a) biología y salud; b) ciencias sociales (a través de la información que brindan las redes sociales) y c) finanzas y negocios. En la sección cuarta se enumeran algunas recomendaciones de instituciones internacionales e investigadores y por último, en la quinta sección se presentan las conclusiones de estas reflexiones.

2. Conceptos básicos para el tratamiento ético de los datos

Como ya se señalara, los tres conceptos básicos en el tratamiento ético de datos son: privacidad, confidencialidad y anonimato. Al trabajar con grandes bases de datos, estos conceptos han cobrado nuevas dimensiones y deben, por lo tanto, ser cuidadosamente evaluados en toda nueva planificación para analizar información.

La privacidad se refiere principalmente a la persona, mientras la confidencialidad atañe a la información. Según Sieber (2001) “el respeto por la privacidad significa darles a las personas la privacidad que les gustaría tener, y no las invasiones a su privacidad que habitualmente se les imponen” (pág. 10, traducción propia). La confidencialidad, a diferencia de la privacidad, se refiere al resguardo de la información brindada por los y las participantes (Santi, 2016).

La naturaleza y el entendimiento de privacidad se vuelven confusos en el contexto de investigaciones de *Big-Data*. Por lo tanto, se debe asegurar que será respetada y protegida en este nuevo dominio lo cual se vuelve un gran desafío. Zimmer (2015) ha informado que numerosos estudios indican que el usuario promedio de internet tiene un conocimiento incompleto de cómo sus actividades se rastrean rutinariamente y, por ende ignora que su privacidad está siendo vulnerada relacionada con las visitas de sus sitios. Por ello, no es claro si el usuario de internet realmente entiende cuándo y cómo su actividad es rastreada y monitoreada. El vacío conceptual que existe sobre privacidad y la definición de información personal identificable en las investigaciones a partir de *Big-Data* es inevitable y conduce a tener vacíos similares cuando es necesario el consentimiento informado. En la información de perfiles en las redes sociales, generalmente se argumenta que no es necesario un consentimiento informado específico debido a que la información disponible es pública. Sin embargo, habría que conocer si el usuario realmente entiende las condiciones técnicas bajo las cuales se muestra información visible en una plataforma social, o si saben que sus datos pueden ser recolectados con propósitos de investigación en vez de aparecer en las pantallas para la visualización breve de sus amigos/as o seguidores/as.

En la última década, investigaciones sobre privacidad de la información de usuarios

de redes sociales ha mostrado, a menudo, que usuarios ordinarios ponen poca atención en las políticas de privacidad y que, cuando se consideran estas políticas, es difícil para el común de la gente entender su significado e implicaciones prácticas (Diesner y Chin, 2015). El concepto de privacidad necesita que se dedique mucho tiempo para enseñar a las personas a comprender que sus datos son valiosos y que una vez que están *on line*, éstos se perpetúan para siempre (Jetten y Sharon, 2015). Sin transparencia en la práctica, los usuarios podrían retacear los datos que comparten, especialmente los datos de salud, porque no entienden el valor colectivo que tienen estos datos y la utilidad que otros le pueden dar.

El dominio creciente de investigaciones con *Big-Data*, también conduce al vacío conceptual de cómo se define cuándo realmente existe un daño y si son las personas en sí mismas las merecedoras de protección. Se debe tener en cuenta que a veces los daños no surgen inmediatamente (Zimmer, 2015). Según este autor, el vacío conceptual no se da únicamente en las investigaciones, sino que también en las leyes internacionales y las regulaciones acerca de la recolección y uso de datos personales que varía según la definición de daño. Por ejemplo, en Canadá y la Unión Europea aceptan regulaciones con enfoque muy paternalista de la política de protección de los datos tendientes a preservar los derechos humanos fundamentales de sus ciudadanos/as, a través de acciones gubernamentales preventivas, creyendo que el usuario debe mantener el control sobre la información para preservar su dignidad y autonomía. En contraste, en los EEUU se basan en el supuesto que la recolección de gran cantidad de datos y su utilización son aceptables y beneficiosas, mientras que el límite se debería imponer sólo después que haya ocurrido algún daño tangible.

Hasta ahora los principios de protección de datos no debían aplicarse a información anónima (inclusive con fines estadísticos o de investigación). Sin embargo, con métodos adecuados la tecnología ha permitido reidentificar grandes bases de datos de salud anónimos. Por ejemplo, la AEDP (Agencia Española de Protección de datos) ha reconocido que no es posible considerar que los procesos del anonimato garanticen en un cien por ciento la no re-identificación de las personas, aún con buenas prácticas para preservar el anonimato (García León, 2018). En esta misma línea de pensamiento, Naser (2015) expresa que, dada la potencia de las actuales técnicas computacionales, es imposible que cualquier conjunto de *Big-Data* pueda genuinamente ser anónimo, aún si éste no tiene identificadores específicos unidos a él.

Si bien muchos datos se recogen para un único propósito, pueden ser agregados y usados de múltiples maneras y para muchos otros fines. Esto implica que la procedencia de los datos, que es la base de muchos análisis con *Big-Data*, pueda ser sospechosa ya que los datos no fueron explícitamente provistos para el análisis propuesto. En estos casos, la información que es de los sujetos, se utiliza con fines que nunca imaginaron, aprobaron, entendieron o consintieron. Se debería decir a las personas que aportaron sus datos cómo éstos serán usados. Las regulaciones deberían proveer considerandos para los riesgos y beneficios. Los posibles daños a los sujetos que se incluirían en este tipo de investigación, no sólo son los

concernientes a las tradicionales violaciones de confidencialidad, sino también a la más fundamental condición concerniente a la relación entre información y libertad humana (Naser, 2015).

3. Grandes áreas de investigación que utilizan Big-Data

En este trabajo se hace referencia a tres grandes áreas que más frecuentemente realizan investigaciones usando como fuente *Big-Data* con datos personales.

3.1. Biología y salud

Las investigaciones en salud usando información secundaria están creciendo. Son investigaciones que agregan información sobre pacientes, incluyendo condiciones físicas, información genética, respuesta a tratamientos y resultados. Estas investigaciones muestran a los investigadores un mundo real en una foto, a un nivel de población amplio en una forma que no es posible con los tradicionales ensayos clínicos. Los datos, desde un contexto clínico, son retroalimentados combinando una base de datos y un círculo de retroalimentación continua, que iterativamente ayuda a mejorar diversos resultados clínicos y de salud. La nueva forma de investigación, a primera vista, es una nueva iniciativa política (Konnoth, 2015).

Estas nuevas investigaciones son importantes y muy extendidas, pero requieren de una amplia cantidad de información. Por ejemplo, reunir datos ha permitido investigaciones para identificar mutaciones genéticas que presagian alto riesgo de contraer cáncer de mama o mal de Alzheimer. También se utilizan para investigaciones de usos no clínicos, por ejemplo para ayudar a desarrollar medidas de calidad en salud necesarias, que permitan identificar áreas con objetivo de reducir costos. También pueden respaldar otros mecanismos tales como los ensayos clínicos.

Como se ve, los frutos de estas nuevas investigaciones son muchos pero imponen cargas sobre las personas. La involuntaria liberación de información en salud puede objetivamente causar daños sobre personas, si es usada inapropiadamente, yendo desde discriminación en el empleo hasta pérdida de seguridad de su reputación. Aunque estos daños eventualmente no ocurrieran, la incomodidad psicológica que se produce por temer su utilización inapropiada o si no se estima el riesgo de éste u otros daños, es en sí mismo otro daño. Finalmente, aún si no se teme que la información sea mal usada, las personas tienen una experiencia de intromisión por no respeto a su privacidad y daño de su autonomía (Konnoth, 2015). La persona se vuelve transparente. También se debe realizar una distinción entre los datos recogidos con el propósito de investigar, de los que no tienen dicho fin. Además, es necesario un método sistemático para definir datos sensibles.

Tradicionalmente, se aceptaban internacionalmente características que tienen las guías para profesionales de la salud donde dicen qué deberían hacer y qué no cuando se manejan datos de salud. Esto era necesario porque los datos de salud tenían habitualmente protecciones especiales garantizadas. Sin embargo, en la era de los avances rápidos de la tecnología, características aceptadas previamente no son suficientemente extensas por tres razones. Primero, las innovaciones superan las características en sí mismas, muchas de las cuales se reflejan en la forma de recolección de datos de décadas anteriores. Segundo, los datos de salud están siendo

recogidos fuera del tradicional marco de la asistencia sanitaria. Tercero, los datos son compartidos, en una tercera parte, no sólo para investigación, sino también para beneficios comerciales (Jetten y Sharons, 2015). Estos mismos autores, se realizan dos preguntas que son pertinentes en la asistencia en salud: 1) ¿es ético controlar o influenciar el comportamiento del usuario en salud mediante *Big-Data*? y 2) ¿debería la información potencialmente salvar vidas al ser compartida con pacientes que han consentido compartir datos pero para un propósito no relacionado? Agregan una tercera a tener en cuenta: ¿se debería comunicar un hallazgo inesperado en los datos del sujeto o se debería permitir que éste tenga la opción de elegir si quiere conocer estos hallazgos accidentales?

En cuanto al área genética, Lima (2008) sostiene que:

El fenómeno de Big Data combinado con los avances e innovaciones propios de la Era Genómica ha generado una red de alcance mundial que despierta retos globales.

Uno de los efectos más inmediatos de este fenómeno social ha sido la aparición de una narrativa en relación al avance y a la manipulación genética que pone en cuestión importantes aspectos sociales (pág. 11).

La misma autora expresa que “Demás está decir que en muchos países existe un gran vacío normativo, donde no se ha logrado arribar a una reglamentación, o bien, estas investigaciones han sido consideradas como experimentales y no terapéuticas” (pág. 12).

La genética es una disciplina que irrumpe permanentemente con nuevas herramientas que permiten entender aspectos del genoma y su funcionamiento pero, como todo en ciencia, nuevas situaciones llevan a nuevas preguntas. La edición de genes es la herramienta que hoy en día permite editar al ADN (Ácido desoxirribonucleico). La utilización de esta metodología tiene gran posibilidad de convertirse en herramienta terapéutica para distintas enfermedades humanas. Los avances metodológicos de gran impacto en la sociedad necesitan ser mirados críticamente, comprendidos en su alcance, examinados desde la utilidad que brindan y contar con la posibilidad de interrogarse, en el caso de la especie humana, sobre su utilidad y sobre todo en sus limitaciones éticas cuando se manejan grandes bases de datos de genomas.

3.2. Ciencias sociales

A partir de la información que brindan las redes sociales han surgido nuevas formas de investigaciones en ciencias sociales, las cuales también deben ser contempladas desde un punto de vista ético. Por ello, definir prácticas éticas para las investigaciones que utilizan las redes sociales y digitales es un desafío actual.

Con 2,5 trillones de datos creados y registrados todos los días, la gente está definiendo cómo viajan, comen, navegan en internet y viven sus vidas. Se está en el medio de una “revolución de datos”, donde las personas y las organizaciones pueden almacenar y analizar cantidades masivas de información. El manejo de datos, sorpresivamente, ahora permite descubrir e innovar información teniendo un potencial fundamental para, por ejemplo, cambiar sociedades (Erickson, Harris y Lee, 2018). Estos autores agregan: trabajar con grandes bases de datos, nuevas herramientas analíticas y métodos permitidos por los datos es prometedor para

conseguir muchos servicios para la sociedad. Sin embargo, tecnologías de manejo de datos también desafían el supuesto fundamental acerca de sobre qué muestra de la sociedad éstos se construyen.

Los comités éticos suelen poner especial atención a la selección de los sujetos de la investigación. Esta es un área sub-investigada en investigaciones con datos de redes sociales y digitales. Con la participación *on line*, los sujetos se seleccionan a sí mismos y la comunidad *on line* puede tener características distintas de la población general. También puede ser difícil para el investigador poder decir si sus participantes pertenecen a alguna población vulnerable. Esta reflexión es necesaria, para advertir los potenciales sesgos en los estudios generados con datos *on line* (Shilton, 2016).

Respecto a la ética en ciencias sociales Santi (2016), expresa:

Si bien en los últimos años ha habido numerosos debates en torno a diversos aspectos éticos de las investigaciones sociales, no obstante, en tanto campo de estudio no ha tenido un desarrollo sistemático a nivel global. De hecho, no existen actualmente lineamientos éticos con respecto a la realización de estas investigaciones que sean aceptados internacionalmente y que puedan ser aplicados de forma general” (pág. 7). “quisiera esbozar un nuevo desafío ético en cuanto a la protección de la privacidad y al resguardo de la información. Esta problemática surge a raíz de las nuevas formas de hacer investigación a través del uso de internet, o tomando información de los foros de debates, en los sitios de chat o en las redes sociales. Un punto a destacar en este tema es que si bien el uso de internet ha expandido las posibilidades de investigación del comportamiento e interacción humanas, también ha planteado o renovado problemas éticos (pág. 17).

Cuando se manejan técnicas flexibles en un proyecto, los investigadores pueden, algunas veces, tomar un enfoque pragmático para trabajar con tablas de datos sociales. Un ejemplo podría ser considerar tales datos como un medio para identificar patrones generales de interacción social y estructura, tanto como la dinámica subyacente del sistema técnico-social. Conceptualizar los datos sociales como una herramienta para obtener conocimiento generalizado acerca de la sociedad, grupo o comunidades como un todo (lo opuesto a identificar personalmente a sus miembros), puede explicar por qué algunos investigadores son inconscientes acerca de la ética e implicaciones sociales de sus trabajos (Diesner y Chin, 2015).

Son bien conocidos los problemas ante el Congreso y la justicia de EEUU del dueño de Facebook, Mark Zuckerberg, sobre su autorización que permitió a una firma de ingeniería social acceso a detalles personales de usuarios de su empresa. Como consecuencia de ello es mundialmente conocida la interferencia que produjo la información que posee Cambridge Analytica en múltiples elecciones de distintos países, entre ellos la elección sobre el Brexit en el Reino Unido (Caldwalladr, 2018). La Unión Europea votó a favor de realizar su propia Regulación de Protección de Datos General (GDPR, General Data Protection Regulation), que proporciona guías para compañías de tecnología y comunicación en las 28 naciones miembro cuando

se trata de datos personales de usuarios. Esto ha sido bajo la amenaza de fuertes multas si las compañías tecnológicas y de datos no pueden proveer transparencia a los usuarios sobre cómo sus datos son recolectados y usados (Nichols, 2018).

3.3. Negocios y finanzas

La importancia que tiene el Big Data en los negocios y en las finanzas es reconocida por varios autores, entre los que se pueden distinguir lo que expresa Gadi (2015): “Hoy prácticamente todo el mundo entiende que el Big Data está revolucionando el ámbito de la empresa y de la economía”; mientras Miranda (2017) dice:

El conocimiento en tiempo real del cliente digital aprovechando las fuentes de información no estructural harán del Big Data una herramienta imprescindible en el mundo financiero (...) el sector financiero es líder en todo lo referente a propiciar desarrollos tecnológicos relacionados con la toma de decisiones apoyados en la ingente cantidad de datos que proporciona el *Big Data*.

Nieuwbourg (2015) agrega:

La fórmula ideal para representar Big Data sería (Volumén + Velocidad + Variedad) x Visualización = Valor (...) si hay un área en la que el valor se puede medir y donde es importante, es en la detección de fraude del sector financiero.

De forma sistemática se pueden resumir las ventajas de la utilización de *Big Data* en que permite: 1) segmentar clientes, 2) análisis y gestión de riesgo, 3) detección de fraude y de blanqueo de capitales y 4) la retención de clientes.

Las redes sociales son usadas, en muchas ocasiones, por entidades financieras como los bancos y también por compañías de seguro. Por ejemplo, los bancos a través de cruce de información y algoritmos apropiados pueden detectar fraudes con tarjetas de crédito. Las compañías de seguros pueden determinar circunstancias de falta de precaución de sus clientes ante robos en su propiedad u otro tipo de siniestro. Si bien la utilización de información de las redes sociales puede ser muy útil para obtener más ganancias a este tipo de compañías e incluso, a veces, beneficios para sus clientes es indispensable preguntarse cómo se cubre el aspecto ético de este tipo de investigaciones, porque de hecho se está violando la privacidad de las personas y se están utilizando los datos para otro fin diferente que para el que fueron expuestos en las redes sociales.

Los perfiles y la focalización de los usuarios a través de un análisis normativo pueden dar lugar al manejo del comportamiento del consumidor. En este caso, el consumidor actuará bajo una oferta personalizada más refinada, lo que puede llevar a que un conjunto de consumidores no piensen en otras alternativas del mercado porque están menos conscientes de ellas. Eliminando opciones de las consideraciones del consumidor se interfiere en el principio de libre elección. Esta práctica puede ser comercialmente interesante, pero éticamente es una pendiente resbaladiza tal como señalan Jetten y Sharon (2015). El consumidor debería tener habilidad para poder revisar los datos que son coleccionados sobre él y tener la posibilidad de retirarse en cualquier momento, además de tener claro el beneficio de compartir sus datos.

Con respecto al mecanismo en los bancos Gadi (2015) dice:

El negocio de los bancos se basa en captar ahorro y prestar dinero y cobran por estos servicios un interés que, en algunos casos, viene determinado por el grado de solvencia del prestatario. Toda la vida se ha hecho así y los bancos tienen muy buenos sistemas para controlar y determinar la capacidad de pago o devolución de deudas de sus clientes. Sin embargo esa información es solo una décima parte del volumen de información total que hay en la red. Una información que es pública y a partir de la cual se pueden extraer múltiples variables que también podrían determinar nuestra credibilidad, solvencia o capacidad de pago. A través del rastro que dejamos en internet (redes sociales, teléfonos, grupos de interés, etc.) se pueden determinar múltiples variables que podemos poner al servicio de nuestro negocio. Son los proxys (información alternativa) que, si son bien trabajados, eliminarían la aparente ventaja competitiva que tiene la banca u otros sistemas de pago.

Basándose en que la información en las redes es pública, en el área de las finanzas y negocios es donde menos se habla de las precauciones éticas que se deben tener en este tipo de investigaciones y algunas de ellas no son muy precisas como lo que expresa Miranda (2017) al respecto:

será preciso que las entidades financieras sean capaces de captar de sus clientes mayores volúmenes de información, lo que les permitirá afinar aún más en las ofertas de productos. Obviamente esto debe realizarse de forma cuidadosa, de forma que el cliente no perciba que se está invadiendo su intimidad de forma abusiva y que los datos personales que está aportando realmente sirven para mejorar la calidad de los productos que se le ofrecen.

Por todo lo enunciado, éste área necesita de una mayor discusión de los recaudos éticos que se deben tomar al trabajar con datos brindados sobre las personas por las redes sociales.

4. Recomendaciones de instituciones e investigadores

Las recomendaciones sobre cuestiones éticas en investigaciones tienen que ver con la cultura y característica de los distintos países; no obstante, se enumerarán algunas de las principales que han realizado organismos gubernamentales e investigadores. Como expresa Santi (2016) “Al ser una problemática relativamente nueva —la de investigar en internet— solo un reducido número de organismos e instituciones han elaborado lineamientos éticos generales que podrían sugerirse para hacer una investigación de este tipo” (pág.18).

Como se mencionó anteriormente, en EEUU la ética en investigación ha sido guiada por el Belmont Report (IBR), con foco en el respeto a las personas, beneficios y justicia (Office of the Secretary of the National Commission for the Protection of Human Subjects of Biomedical Research, 1979). Este documento fue muy importante y luego en el 2012 el Department of Homeland Security emitió el Menlo Report, presentando los principios éticos que guían la información y comunicación en investigación tecnológica. Sólo se agregó el principio de Respeto a la Ley de Interés Público. Estos principios éticos aplican bien para investigaciones con *Big-*

Data.

Sin embargo, hoy día uno de los problemas es que el IBR y su ampliación fueron sobrepasados por los nuevos adelantos tecnológicos (Garfinkel, 2015). Este último autor también hace algunas recomendaciones adicionales que se deberían incluir tales como: 1) separar la aprobación para recoger los datos de la de su análisis; 2) requerir entrenamiento adicional para datos muy sensibles; 3) permitir la codificación como una alternativa a la destrucción de datos; 4) especificar el estándar de seguridad; 5) desarrollar estándares para los datos originarios y datos compartidos y 6) considerar la notificación comunitaria como una alternativa al consentimiento individual y, finalmente, constituir comités de comunidades de consentimiento. Estas recomendaciones pueden paliar en alguna medida lo expresado por Shilton (2016) sobre que el respeto a las personas ha sido ampliamente interpretado en las revisiones éticas como un mandato para obtener el consentimiento informado de los participantes cuando se recogen datos privados. Datos de redes sociales y disponibilidad abierta de datos digitales pueden ser interpretados como públicos; sin embargo, conseguir el consentimiento informado en la escala en que se usan estos datos puede ser dificultoso o imposible.

Las pautas éticas del Open Data Institute (organización sin fines de lucro cuya sede central está en el Reino Unido y tiene como objetivo la difusión de datos abiertos en internet, a través del Esquema de la Ética de Datos) proveen una guía de los distintos aspectos que se deben plantear cuando se realizan investigaciones. Las mismas se pueden resumir de la siguiente manera: 1) descripción de la fuente de datos que comprende: ¿cuáles son sus características principales?, ¿quiénes tienen permiso y derechos sobre ellos y con quiénes se comparten?, ¿qué limitaciones pueden tener en la actualidad?, ¿qué políticas y legislaciones le son de aplicación); 2) propósito del uso de los datos, ¿entiende la gente cuál es el propósito?, ¿quiénes pueden verse positivamente o negativamente afectados?, ¿cómo se podría minimizar el impacto negativo?, ¿cómo podrían los afectados interactuar con la organización? y 3) gestión de riesgo, ¿se están comunicando los riesgos adecuadamente?, ¿cuál es la política de revisión de riesgo?, ¿qué acciones están previstas?.

La Comisión Europea a través de sus “Pautas para la protección de los datos y privacidad”, cuya última reforma se produjo en el 2018, dice: “Las normas más estrictas en materia de protección de datos implican que las personas tienen más control sobre sus datos personales y que las empresas se benefician de igualdad de condiciones”. En su página web brinda un apartado sobre “Derechos de los ciudadanos”, donde explica estos derechos, y “Normas para empresas y organizaciones. Aplicación del reglamento general de protección de datos (RGPD), obligaciones, solicitudes de personas y cumplimiento”, donde se expresan estas normas y las sanciones en caso de no cumplimiento.

García León (2018), apunta algunas notas, derivadas del RGPD, acerca de la investigación en salud a partir de *Big-Data*, de las cuales se pueden resaltar los principios que deben tener en cuenta los investigadores.

Por otro lado, Richards y Hartzog (2015) expresan que últimamente la paranoia con *Big-Data* es justificada pero no productiva. Se debería buscar en el futuro una forma

que permita explorar algunos de sus potenciales beneficios y mientras tanto eliminar algunos de sus riesgos obvios. Crean que la confianza y la ley deberían ser el norte de la ética en *Big-Data* y que para conseguir confianza, las organizaciones deberían comprometerse a proteger los datos, a ser discretos y honestos y lo que estos autores llaman *loyalty*, esto es, que las empresas no se enriquezcan a expensas de la confianza de sus usuarios.

5. Conclusiones

En este trabajo se ha realizado una revisión de trabajos de los últimos años sobre ética en investigación cuando se trabaja con *Big-Data* que incluya datos personales. De ello se puede concluir, como lo expresan varios investigadores e instituciones, que los conceptos de confidencialidad, privacidad y anonimato deben ser revisados ante el advenimiento de nuevas tecnologías y metodologías que tratan estas grandes bases de datos.

Aún esta discusión sigue vigente y si bien algunas instituciones internacionales e investigadores han confeccionado recomendaciones, el debate está abierto y se reconoce la difícil tarea de conciliar los beneficios que pueden brindar estas investigaciones con el mantener los principios éticos, en especial la privacidad y no manipulación de los participantes. Lo que algunos autores consideran que se enmarcaría dentro de los derechos humanos.

También se plantea que la definición de los conceptos éticos que deben primar en este tipo de investigaciones depende de la cultura y leyes de cada país. Por ello, el tema ético de investigaciones con *Big-Data* con datos personales, en cualquier área de la ciencia, debería discutirse en forma multidisciplinaria, tanto en medios académicos, gubernamentales y profesionales, como así también ser transmitida y planteada a estudiantes universitarios en sus respectivas curriculas, junto con una divulgación accesible y clara para la población general.

Referencias bibliográficas

- Ayers, R. (2018). New legal and ethical challenger of big-data. *Smart Data Collective*. Recuperado de <https://www.smartdatacollective.com/new-legal-ethical-challenges-of-big-data/>.
- Cadwalladr, C. (2018). La Guerra de los datos. *La pupila insomne*. Recuperado de: <https://www.la-pupila-insomne.wordpress.com/2018/03/20/la-guerra-de-los-datos-por-Carole-Cadwalladr/>.
- Clifton, C. (2015). Ethics Review Process as a Foundation for Ethics Thinking. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Croll, A. (2012). Big Data in our generation's civil rights issue, and we don't know it. *O'Reilly y Radar*. Recuperado de <http://radar-oreilly.com/2012/08/big-data-in-our-generation's-civil-rights-issue-and-we-don't-know-it>.
- Diesner, J. y Chin, C. (2015). Usable Ethics. Practical Consideration for Responsibly Conducting Research with Social Trace Data. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.

- Erickson, L. C., Harris, N. E. y Lee, M. M. (2018). It's time to talk about ethics. *Oracle+Datasciencia*. Recuperado de <https://www.datascience.com/blog/data-ethics-for-data-scientists>.
- Gadi, M. (2015). Las nuevas finanzas se llaman Big Data. *Cinco Días. El País*. Recuperado de https://cincodias.elpais.com/cincodias/2015/08/11/empresas/1439314515_601272.html
- García Leon, J. (2018). Ética, investigación, bases de datos. *Javier García León*. Recuperado de <https://javiergarcialeon.net/2018/11/12/etica-investigacion-bases-de-datos/>.
- Garfinkel, S. L. (2015). Beyond IRBs. Designing Ethical Review Processes for Big Data Research. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Hirsch, D. D. y King, J. H. (2015). Big Data Sustainability an Environmental Management System Analogy. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Jetten, L. y Sharons, S. (2015). Selected Issues Concerning the Ethical use of Big Data Health Analytics. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Konnoth, C. (2015). Classification Standars for Health Information: Ethical and Practical Approaches. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Lima, N. S. (2018). CRISPR/Cas9: *Reflexiones Bioéticas sobre las modificaciones genómicas*. *Journal of Basic and Applied Genetics*. Vol XXVIV(1), 9-15.
- Miranda, M. S. (2017). Big Data herramienta clave del sector financiero. *Forbes, México*. Recuperado de <https://www.forbes.com.mx/big-data-herramienta-clave-del-sector-financiero/>
- Naser, C. (2015). The IRB Sledge-Hammer, Freedom and Big Data. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Nichols, M. R. (2018). Big Data is so large, it's raising privacy and Ethical issues. *Scienia in society*. Recuperado de <https://www.euroscientist.com/big-data-ethical-issues/>.
- Nieuwbourg, P. (2015). Big Data en las finanzas: desde ya aplicaciones práctica. *Divideo*. Recuperado de https://www.decideo.com/Big-Data-en-las-finanzas-desde-ya-aplicaciones-practicas_a1192.html.
- Olivera, J. (2018). Is big-data a big ethics problem? *SAS Blogs*. Recuperado de <https://blogs.sas.com/content/hiddeninsights/2018/10/16/data-ethics-problem/>.
- Richards, N. y Hartzog, W. (2015). Trusting Big Data Research. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.

- Santi, M. F. (2016). Controversias éticas en torno a la privacidad, la confidencialidad y el anonimato en investigación sociales. *Revista Bioética y derecho*, (37).
- Shilton, K. (2016). Emerging Ethics Norms in Social Media Research. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Sieber, J. (2001). Privacy and Confidentiality: As Related to Human Research in Social and Behavioral Science. En NBAC, *Ethical and Policy Issues in Research Involving Human Participants*, Vol. II, Maryland, EEUU.
- Tene, O. y Polonitsk, J. (2015). Beyond IRBs. Ethical guidance for data research. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Vayena, E., Grasser, U., Wood, A., O'Brian, D. R. y Altman, M. (2015). Toward a New Ethical and Regulatory Framework for Big Data Research. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.
- Vogelberg, K. M. (2019). La ética y la tecnología inseparables para cualquier sector. *Retina, El País*. Recuperado de https://retina.elpais.com/retina/2019/02/12/tendencias/1549974578_923870.html
- Zimmer, M. (2015). Research Ethics in the Big Data Era. Addressing Conceptual Gaps for Researchers and IRBs. *Big Data Ethics*. Recuperado de <https://bigdata.fpf.org/papers/>.