

# Una aproximación interdisciplinaria de Big Data al Primer Nivel de Salud: Aportando al proceso de elaboración de políticas públicas

Juan Pablo Ron  
Universidad Tecnológica Nacional  
Facultad Regional Rosario  
Email: juanpabloron@gmail.com

Santiago Bastian  
Universidad Tecnológica Nacional  
Facultad Regional Rosario

Guillermo Leale  
Universidad Tecnológica Nacional  
Facultad Regional Rosario

Carolina Colombres  
Universidad Nacional de Rosario  
Facultad de Ciencias Médicas

Johanna Gerke  
Superintendencia de Riesgos del Trabajo  
Ministerio de Trabajo, Empleo y Seguridad Social  
Email: johannagerke@gmail.com

**Resumen**—El presente trabajo tiene como objetivo explorar los desafíos de estudiar interdisciplinariamente la relación Salud-Minería de Datos-Big Data en la Argentina. El mismo forma parte de un proyecto de investigación interdisciplinario sobre big data y salud de un equipo conformado por profesionales de las ciencias médicas, las ciencias sociales y la ingeniería en sistemas.

De esta manera, se presenta aquí un primer análisis del sector primario de salud, que se caracteriza por ser un nivel de atención médica de baja complejidad y alta demanda, especialmente en los barrios donde los pacientes toman el primer contacto con el sistema de salud público.

La aplicación de técnicas de big data permite construir una base para la toma de decisiones gubernamentales, identificando y conociendo las problemáticas y necesidades específicas de cada barrio. Actualmente, los datos estructurados recolectados por los Sistemas de Información de Salud (S.I.S.), junto a datos no-estructurados, como textos de historias clínicas, videos, imágenes y registros de sensores son obtenidos en los barrios mediante actividades de prevención, promoción y educación de la salud.

Uno de los retos que se presentan en la dinámica de recolección de datos, es mantener la transparencia en la gestión de los datos y la privacidad de las personas, garantizando su intimidad frente a la revolución informática.

Cabe destacar, que el proceso de recolección de datos y su análisis para la identificación de necesidades y problemáticas particulares es llevado a cabo de manera interdisciplinaria, sabiendo que la construcción del dato lejos está de responder a criterios “objetivistas” y neutrales. Por este motivo, el resultado del análisis ofrecerá la posibilidad de construir indicadores reales para generar acciones para el desarrollo y promoverlas en el marco de políticas públicas.

*Index Terms:* Big Data - Data Mining - Salud - Políticas Públicas

## I. INTRODUCCIÓN

¿Qué puede aportar Big Data al diseño y gestión de las políticas públicas? Un ejemplo ineludible para responder a este interrogante es el conocido virus de la influenza A (H1N1). Durante septiembre del 2009 los centros de control y prevención de enfermedades de Estados Unidos pedían a los médicos que informaran los nuevos casos de “gripe A”, el

problema era que el registro y procesamiento de la información llevaba su tiempo y las unidades de control sólo obtenían resultados cada una o dos semanas. Tiempo más tarde la revista científica Nature publicó una nota donde se explicaba cómo Google podría predecir la zona de propagación de la gripe [1]. Debido a que Google recibe y archiva millones de consultas de usuarios, podía identificar a los afectados por el virus H1N1 según las búsquedas realizadas en internet, obteniendo resultados en tiempo real. En este caso, con el objetivo de predecir y prevenir la propagación de la pandemia, se obtiene un valor significativo basado en el análisis de los *Grandes Datos* (en inglés *Big Data*).

El objetivo principal de este trabajo es explorar los desafíos de estudiar interdisciplinariamente la relación Salud-Minería de Datos-Big Data en la Argentina; con el propósito que sirvan de insumo para política pública y la construcción de acciones específicas.

El mismo forma parte de un proyecto mayor para el que hemos conformado un equipo interdisciplinario e interinstitucional entre docentes e investigadores de las siguientes universidades: Universidad Tecnológica Nacional de Rosario, Universidad Abierta Interamericana de Rosario, Universidad Nacional de Rosario, Universidad Nacional de Quilmes y la Universidad Nacional de San Juan. Con esta sinergia de fuerzas provenientes de distintas disciplinas que cruzan lo social, la ingeniería en sistemas y la informática, se está realizando un proyecto de investigación con el fin de lograr una perspectiva que integre la aplicación de la Minería en grandes Bases de Datos para efectuar aportes significativos desde el ámbito académico en la mejora de las condiciones referidas a la salud en el ámbito público.

En el presente artículo partimos de una experiencia en un centro de Atención Primaria de la Salud (APS). Luego se aborda el concepto de big data y su potencial inclusión en la gestión de políticas públicas, explorando la participación ciudadana en el ámbito de salud, el concepto de privacidad de

datos personales y las leyes que rigen en Argentina.

A continuación, se expone la necesidad de trabajar en equipos interdisciplinarios capaces de abordar las diferentes dimensiones que implica el análisis del primer nivel de salud pública. Se presenta una prospectiva sobre el despliegue de las siguientes fases de la investigación sobre la que se erige el presente artículo, definiendo el trabajo a futuro y estableciendo objetivos claros para llevar en la práctica del proyecto.

Finalmente, en las conclusiones nos aproximamos a una síntesis analítica sobre las posibilidades que podría otorgar la intersección de técnicas big data para mejorar el servicio y atención de salud.

## II. RELEVAMIENTO DE LAS CIENCIAS MÉDICAS EN UN CENTRO DE ATENCIÓN PRIMARIA DE LA SALUD

El presente trabajo fue inspirado por una investigación realizada por profesionales médicos en un centro de Atención Primaria de la Salud (APS), en una ciudad de la Argentina durante el año 2015. La hipótesis de la misma fue que los motivos de consultas se relacionan con las características ambientales, geográficas, socioeconómicas y culturales, según el acceso a la educación, al agua potable y a la cobertura de saneamiento.

Para reflexionar acerca del concepto de salud y APS partimos de la definición de Pedro Luis Castellanos, el cual plantea que los procesos de salud -enfermedad son una expresión de las condiciones de vida de los diferentes grupos poblacionales y que la explicación de esos procesos se rige por la forma de reproducción social de las condiciones objetivas de existencia. Por lo tanto, el abordaje de los problemas de salud en este espacio se debe hacer teniendo en cuenta que surgen de la interrelación de todos los aspectos del proceso de reproducción social (biológico, ecológico, cognitivo, de inserción en el proceso productivo) que componen las condiciones de vida de los grupos poblacionales y cualquier hecho que involucre a uno está afectando a todos.

El escenario del caso en estudio se presenta en este nivel de atención, de baja complejidad y alta demanda, donde las personas toman el primer contacto con el sistema de salud público. La recolección de los datos estuvo a cargo de integrantes del equipo de salud (enfermeros, médicos, agentes sanitarios, trabajadora social, odontólogas, personal administrativo, psicólogos), vecinos e instituciones que forman parte de la comunidad. Se utilizó como guía el mapa que describe geográficamente la zona donde está ubicado en centro de salud y el contexto que rodea al mismo para analizar las características ambientales, geográficas, socioeconómicas y culturales. Las consultas médicas que se evaluaron fueron realizadas por los ocho profesionales que trabajan en el Centro, los mismos son: tres generalistas, dos pediatras, una médica ginecóloga, un médico obstetra y una médica clínica. Aquí se presentan sólo datos de las variables sexo y motivo de consulta.

Entre los principales resultados, se observa en la Figura 1, que el principal motivo de consulta médica fue Control de Salud (30%), seguido por Bronquitis Obstructiva Permanente (14%) e Infección respiratoria (13%). Los tres motivos

coinciden con las especialidades médicas que recibieron más demanda (pediatría y generalista).

Adicionalmente, cabe aclarar que las afecciones respiratorias (Bronquitis Obstructiva Permanente e Infección Respiratoria) alcanzan el 27% de los casos analizados ya que el relevamiento se realizó en período invernal, época habitual en la que aumenta la prevalencia de estas enfermedades en adultos y especialmente en niños.

En la Figura 2 se observa que tres de cada cuatro consultas fue realizada por mujeres. La pronunciada tendencia de consultas del sexo femenino, puede atribuirse a los controles ginecológicos que se realizan, tanto control ginecológico de rutina (Incluido en la categoría Control de Salud 30%), Planificación Familiar (4%), Control de Embarazo y Puerperio (6%) y Papanicolau (10%).

Abordar los problemas de salud del sistema de atención primario, permite identificar las siguientes variables a nivel ecológico (condiciones naturales, saneamiento ambiental, características del barrio), a nivel de las formas de conducta (organización, participación, educación) y a nivel de las relaciones económicas (proceso de trabajo, participación en la distribución, en el consumo, estado y evolución de la vivienda) brindando información para mejorar el servicio sobre el desarrollo de las personas o bien detectar la necesidad de cambiar las reglas generales del sistema. El tipo de acciones que se derivan de este abordaje en particular se debe organizar en planes y programas de salud, abriendo mayores posibilidades para redefinir la prevención y promoción de la salud, como acciones sociales dirigidas a promover la transformación de la calidad de vida de todos los sectores de la población.

Conforme al enfoque metodológico de la investigación-acción, según la información relevada y la perspectiva de aplicaciones de técnicas de big data, se plantea como objetivo principal el desarrollo de herramientas informáticas innovadoras capaces de contribuir como insumo de políticas públicas que den soporte a la estrategia y a la toma de decisiones para el desarrollo de planes y programas en el ámbito de salud, procesando la información recolectada de la participación ciudadana en el contexto del primer nivel de salud pública.

## III. BIG DATA, SISTEMA DE INFORMACIÓN DE SALUD, DATOS PERSONALES Y DERECHOS DEL PACIENTE

El concepto big data refiere a conjuntos de datos tan grandes como diversos, que rompen la infraestructura de tecnologías de información tradicionales utilizadas para capturar, almacenar, gestionar y procesar los datos dentro de un tiempo tolerable, con el fin de obtener resultados de valor para el usuario interesado [2].

Por lo que se refiere al ámbito de la atención sanitaria, a diario se capturan petabytes de datos en los servicios asistenciales —un petabyte (PB) equivale a 1015 bytes = 1.000.000.000.000.000 de bytes—. Los Sistemas de Información de Salud (S.I.S.) en Argentina, se conforman principalmente como se muestran en el Cuadro I [3], [4]:

Hoy en día, los S.I.S. están aumentando el tamaño de los registros de salud disponibles electrónicamente. Simultánea-

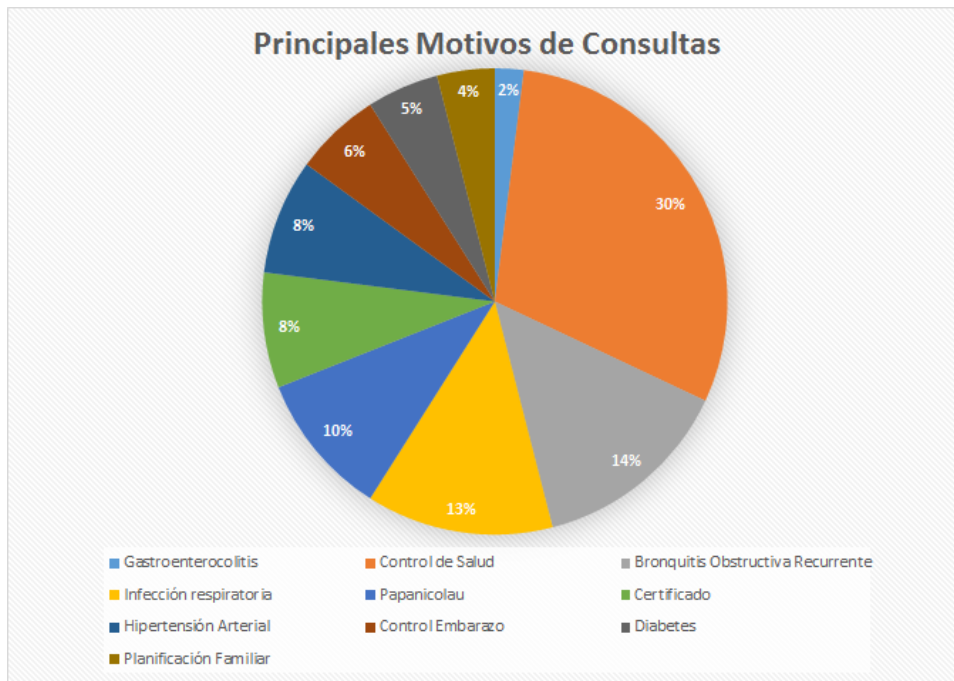


Figura 1. Distribución de frecuencias de los principales motivos de consulta médica en el centro de salud seleccionado entre los meses de junio y agosto de 2015 Argentina

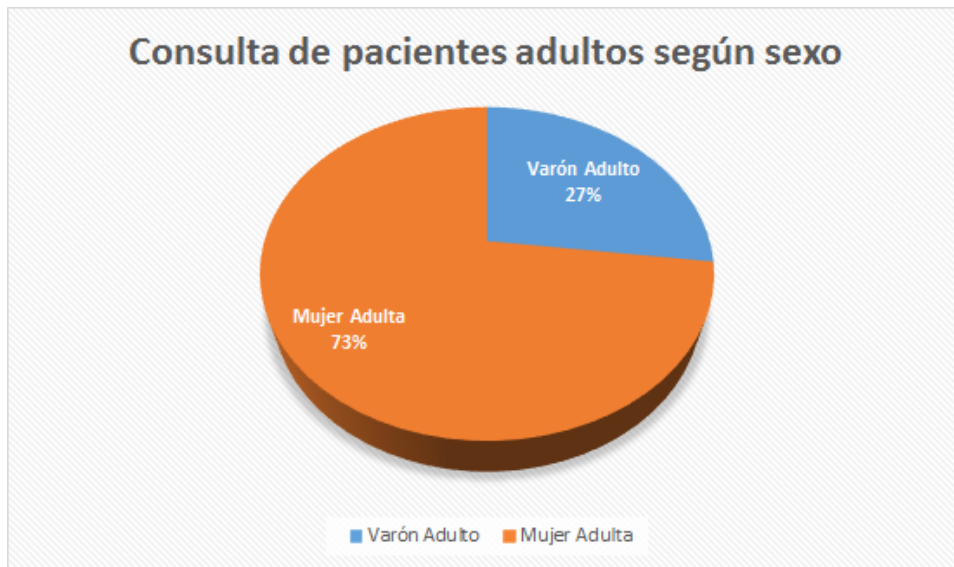


Figura 2. Distribución de consultas de pacientes adultos por sexo en el centro de salud seleccionado, entre los meses de junio y agosto de 2015, Argentina

mente, se han logrado avances en el Análisis Clínico, en inglés Clinical Analytics, con técnicas para analizar grandes cantidades de datos y recoger nuevos conocimientos de ese análisis [11]. Como resultado, hay una serie de opciones para utilizar big data, reducir los costos de la asistencia sanitaria, optimizar el tratamiento de enfermedades, brindar soporte a la toma de decisiones sobre el sistema de asistencia de salud en centros de primer nivel.

Como resultado de la implementación de estas técnicas de Clinical Analytics sobre grandes bases de datos, puede

obtenerse:

- Información detallada de tratamientos de los pacientes que puede determinar tratamientos más eficaces en pacientes individuales o grupos de pacientes.
- Detección de errores en los tratamientos médicos y mitigación de riesgos sobre pacientes asistidos.
- Evaluación del desempeño de las instituciones de salud.
- Nuevas segmentaciones y modelos predictivos en función de los perfiles de los pacientes.
- Automatización de la administración financiera y conta-

Cuadro I  
SISTEMAS DE INFORMACIÓN DE SALUD EN ARGENTINA

Identificación	Descripción	Alcance y Objetivos	Fuente
SES	Sistema Estadístico de Salud	El Sistema Estadístico de Salud se relaciona con el Sistema Estadístico Nacional coordinado por el INDEC. Los contenidos temáticos que aborda se expresan en la producción de estadísticas sobre hechos vitales, condiciones de vida y problemas de salud de la población y sobre disponibilidad y utilización de los recursos.	[5]
SNVS	Sistemas Nacionales de Vigilancia en Salud	El SNVS integra las estrategias de vigilancia clínica, por laboratorios, unidades centinela y de programas específicos, a través de una plataforma montada en Internet que permite la notificación en tiempo real a diferentes niveles (locales, provinciales, nacionales) de distintos eventos bajo vigilancia y el análisis sistemático y rápido de la información disponible que permitan orientar las acciones de salud en cada uno de los niveles involucrados.	[6]
SIVILA	Sistema de Vigilancia de Laboratorio	El SIVILA es el Módulo de Vigilancia por Laboratorio del SNVS. Integra a la vigilancia epidemiológica a laboratorios que procesan muestras provenientes de personas, animales, alimentos y el ambiente. Sus principales objetivos son brindar especificidad a la vigilancia epidemiológica.	[7]
SISA	Sistema Integrado de Información Sanitaria Argentina	El Sistema Integrado de Información Sanitaria Argentino (SISA) es un proyecto de tecnología de la información que da respuesta a las necesidades del Ministerio de Salud de la Nación y los Ministerios Provinciales respecto de la gestión de los datos de sus establecimientos, profesionales, programas o prestaciones para la comunidad. El SISA es un sistema de alcance federal que no reemplaza a los sistemas regionales.	[8]
SMIS	Sistema de Monitoreo de Insumos Sanitarios	Su funcionalidad abarca desde la detección de requerimientos de productos/insumos por parte de una Institución Sanitaria hasta el efectivo ingreso de dichos productos/insumos a los stocks operativos de dicha institución.	[9]
PUCO	Padrón Único Consolidado Operativo	El Padrón Único Consolidado Operativo (PUCO) es un registro de las personas que cuentan con obra social en la Argentina. La búsqueda es individual, por DNI, y la información que devuelve indica si el ciudadano figura o no en PUCO y qué cobertura social posee.	[10]

ble de las instituciones sanitarias.

Por lo tanto, esta nueva capacidad de almacenar, procesar y analizar grandes volúmenes de datos de naturaleza variada en el sector salud podría significar una mejora en la calidad de las asistencias en general, en el incremento de la satisfacción del paciente y en mejoras en la gestión del sistema sanitario.

Sin embargo, también puede vulnerar de forma significativa la privacidad de las personas y su derecho de autodeterminación informativa, sobre todo cuando esta nueva forma de procesar la información tiene el fin de predecir comportamientos o eventos futuros sobre los individuos estudiados. Es por tal motivo, que se prioriza el cuidado de la persona, determinando

y definiendo los límites de la finalidad del dato que garantice la privacidad en el contexto que se obtiene dicha información [12].

En este sentido, en la Argentina existen leyes que respaldan el cuidado de los individuos con respecto a la recolección de datos. Por un lado, la Ley 25.326 sobre de Protección de Datos Personales y la Ley 26.529 de Salud Pública que trata sobre los Derechos del Paciente en su Relación con los Profesionales e Instituciones de la Salud, por el otro. Se establece que para las aplicaciones de Big data realizadas con el fin de brindarle un mejor servicio al titular del dato, no sería necesario el consentimiento, en la medida que no se extralimite de los fines

contractuales. Sin embargo, será necesario el consentimiento previo e informado del titular del dato para el desarrollo de análisis de Big data sobre datos personales en los que la finalidad del tratamiento sea distinta o no compatible con la que motivó su recolección.

En cuanto a la accesibilidad de la información obtenida para posteriores investigaciones, en septiembre del 2016 fue promulgada la Ley 27275 sobre Derecho de acceso a la información pública. Esta tiene por objeto garantizar el efectivo ejercicio del derecho de acceso a la información pública, promover la participación ciudadana y la transparencia de la gestión pública [13].

En este contexto de revolución informática, la participación ciudadana cumple un rol esencial en el aporte de información sobre los determinantes sociales de la salud. Entendemos por estos a las circunstancias en que las personas nacen, crecen, viven, trabajan y envejecen, incluido el sistema de salud. Esas circunstancias son el resultado de la distribución del dinero, el poder y los recursos a nivel mundial, nacional y local. Estos determinantes explican la mayor parte de las inequidades sanitarias, esto es, de las diferencias injustas y evitables observadas en y entre los países en lo que respecta a la situación sanitaria.

#### IV. UNA PROPUESTA INTERDISCIPLINARIA PARA EL DESARROLLO DE LA SALUD PÚBLICA

Las métricas e indicadores de la medicina y la epidemiología se co-construyen a partir de la evolución tecnológica y la dinámica social, haciendo referencia al hecho de compartir objetivos cognitivos comunes, que brindan soporte a la toma de decisiones gubernamentales. La información disponible en Internet, en las redes de telecomunicaciones, en los sistemas de gestión de salud, facilitan el acceso no solo en formatos estructurados y en bases de datos tradicionales, sino también en forma de imágenes, voz, posicionamiento geográfico, representando tipos de datos no estructurados. Este gran volumen de información, debe ser almacenada y procesada de forma distribuida mediante técnicas de big data, obteniendo indicadores que aporten como insumo de política pública.

El proceso de registración de estos grandes datos y su análisis para la identificación de necesidades y problemáticas particulares debe ser llevado a cabo de manera interdisciplinaria, sabiendo que la construcción del dato lejos está de responder a criterios “objetivistas” y neutrales. Entendemos la investigación como un proceso a través del cual se pretende acceder a la esencia del fenómeno de estudio, es decir, a su entendimiento y comprensión; el proceso por medio del cual el investigador expande los datos más allá de la narración descriptiva [14]. Por este motivo, pretendemos que los resultados de los distintos análisis del grupo ofrezcan la posibilidad de construir indicadores reales para generar acciones en el marco de políticas públicas.

El enlace, Salud-Minería de Datos-Big Data, surge de nuestro principal propósito: garantizar el bienestar de las personas

dependientes del sistema de salud público, desarrollando y aumentando su accesibilidad, mejorando el servicio, reduciendo costos y optimizando recursos.

Se define utilizar las técnicas de minería de datos en grandes bases de datos de salud pública. En particular, se propone la construcción de un cluster de máquinas de bajo costo para su uso con tecnologías de Big data como Hadoop [15]. Con ello se prevé el acceso, la recopilación, el almacenamiento, procesamiento y análisis de grandes cantidades de datos que, en particular: a) provengan de repositorios de organismos de salud pública; b) tengan en cuenta regulaciones en cuanto a la sensibilidad y privacidad de los mismos; c) consistan en datos de tipo complejo, abundante y no estructurado y d) sean de interés en cuanto no hayan podido ser analizados por medio de metodologías clásicas de análisis de datos, e) permitan instalar un servicio de monitoreo y vigilancia de la salud.

Para ello se debe contemplar en nuestra país la formación de recursos humanos, que incluya estudiantes en ingeniería en sistemas de información especializados en big data para integrarlos al proceso de la investigación científica, así como profesionales de las ciencias sociales capacitados para utilizar de manera autónoma herramientas de la ingeniería en sistemas de la información.

#### V. CONCLUSIONES

El presente trabajo es un avance de un proyecto de investigación en proceso, de naturaleza interdisciplinaria, que nuclea profesionales del campo de la ingeniería en sistemas y de diversas ramas de las ciencias médicas y sociales.

A partir de los datos relevados por los Sistemas de Información de Salud (S.I.S.) podría recurrirse a la metodología de Data Mining ya que esta permite la manipulación de gran volumen y variedad de datos, encontrando patrones y correlación entre variables de un mismo campo en estudio. Las tareas propias de la fase de minería de datos pueden ser descriptivas, descubrir patrones interesantes o relaciones describiendo los datos o predictivas, clasificar nuevos datos basándose en los anteriormente disponibles [16].

Estas metodologías basadas en big data permiten trabajar con los datos de la totalidad de la población en estudio, sin necesidad de confeccionar muestras poblacionales, obteniendo resultados en tiempo real (como se mencionó en el ejemplo de la Influenza H1N1).

De tal forma, así como la investigación en un solo centro de salud, la aplicación de técnicas de big data permitirían el abordaje interdisciplinario de las problemáticas particulares de cada barrio y sus necesidades, dando soporte a la toma de decisiones gubernamentales, obteniendo como resultado tres objetivos principales: la prevención, promoción y educación de la salud en centros de atención primaria.

Adicionalmente, la administración correcta del universo de los datos completos, permitirá implementar un sistema de monitoreo constante y una corriente de datos continua y masiva sobre los pacientes que, al ser analizados en tiempo real, envíe alertas tempranas para mitigar riesgos. También se destaca la notable mejora en los servicios de vigilancia

de epidemias, detectando nodos de infección y tendencias de dispersión para diezmar con medidas ágiles y precisas la epidemia afrontada.

#### REFERENCIAS

- [1] J. Ginsberg, M. H. Mohebbi, R. S. Patel, L. Brammer, M. S. Smolinski, and L. Brilliant, "Detecting influenza epidemics using search engine query data," *Nature*, vol. 457, no. 7232, pp. 1012–1014, 2009.
- [2] L. J. Aguilar, *Big Data, Análisis de grandes volúmenes de datos en organizaciones*. Alfaomega Grupo Editor, 2016.
- [3] A. Chiappello, "Sistemas de información de salud de la república argentina," *Posgrado en Metodología de la Investigación UNR*, disponible en <http://www.saludcolectiva-unr.com.ar/docs/sc-clase04.pdf>. Fecha de último acceso: Abril 2017.
- [4] "Sistema integrado de información sanitaria argentino," disponible en <https://sis.msal.gov.ar/sisa>. Fecha de último acceso: Abril 2017.
- [5] "Dirección de estadísticas e información de salud," disponible en <http://www.deis.msal.gov.ar/index.php/ses/>. Fecha de último acceso: Abril 2017.
- [6] "Área de vigilancia del ministerio de salud," disponible en <http://www.msal.gov.ar/index.php/home/funciones/area-de-vigilancia>. Fecha de último acceso: Abril 2017.
- [7] "Sistema nacional de vigilancia epidemiológica por laboratorios de argentina," disponible en <http://www.msal.gov.ar/index.php/home/funciones/area-de-vigilancia/sivila>. Fecha de último acceso: Abril 2017.
- [8] "Características del sisa," disponible en [https://sis.msal.gov.ar/sisa/sisadoc/docs/0101/intro\\_caracteristicas.jsp](https://sis.msal.gov.ar/sisa/sisadoc/docs/0101/intro_caracteristicas.jsp). Fecha de último acceso: Abril 2017.
- [9] "Sistema de monitoreo de insumos sanitarios," disponible en <http://www.smis.msal.gov.ar/Help/ManualsPopUp.aspx>. Fecha de último acceso: Abril 2017.
- [10] "Padrón Único consolidado operativo," disponible en [https://sis.msal.gov.ar/sisa/sisadoc/docs/0204/puco\\_intro.jsp](https://sis.msal.gov.ar/sisa/sisadoc/docs/0204/puco_intro.jsp). Fecha de último acceso: Abril 2017.
- [11] D. W. Bates, S. Saria, L. Ohno-Machado, A. Shah, and G. Escobar, "Big data in health care: using analytics to identify and manage high-risk and high-cost patients," *Health Affairs*, vol. 33, no. 7, pp. 1123–1131, 2014.
- [12] J. C. González Allonca, E. Ruiz Martínez, and M. F. Pollo Cattaneo, "Buenas prácticas para la protección de datos personales en ambientes de big data," in *Simposio Argentino de GRANdes DATos (AGRANDA 2016)-JAIIO 45 (Tres de Febrero, 2016)*, 2016.
- [13] "Ley 27.275 derecho de acceso a la información pública," disponible en <http://servicios.infoleg.gob.ar/infolegInternet/verNorma.do?id=265950>. Fecha de último acceso: Abril 2017.
- [14] T. G. Gil and A. C. Arana, "Introducción al análisis de datos en investigación cualitativa: Tipos de análisis y proceso de codificación (ii)," *Nure Investigación*, vol. 7, no. 45, 2010.
- [15] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Networks and Applications*, vol. 19, no. 2, pp. 171–209, 2014.
- [16] J. C. Riquelme Santos, R. Ruiz, and K. Gilbert, "Minería de datos: Conceptos y tendencias," *Inteligencia artificial: Revista Iberoamericana de Inteligencia Artificial*, vol. 10, no. 29, pp. 11–18, 2006.