



## ARTÍCULO ESPECIAL

# Ciencia de datos en salud: desafíos y oportunidades en América Latina

*Health data science: Challenges and opportunities in Latin America*

Juan Manuel Rosa<sup>a,b</sup> , Eliana Ludmila Frutos<sup>c</sup>.

<sup>a</sup> Jefe de Informática Médica, Hospital Alemán. Buenos Aires, Argentina.

<sup>b</sup> Docente en Evaluación e Investigación en Informática en Salud, Maestría en Informática en Salud, Instituto Universitario Hospital Italiano de Buenos Aires. Buenos Aires, Argentina.

<sup>c</sup> Departamento de Informática en Salud, Hospital Italiano de Buenos Aires. Buenos Aires, Argentina.

## INFORMACIÓN DEL ARTÍCULO

### Historia del Artículo:

Recibido: 28 07 2022  
Aceptado: 30 09 2022

### Key words:

Data Science;  
Big Data; Health;  
Health Information  
Interoperability; Latin  
America.

### Palabras clave:

Ciencia de Datos;  
Big Data; Salud;  
Interoperabilidad de la  
Información en Salud;  
América Latina.

## RESUMEN

*La ciencia de datos es una disciplina dedicada a analizar y procesar datos con el propósito de convertirlos en conocimiento accionable. En el área de la salud, la ciencia de datos ha tenido repercusión en el último tiempo, principalmente debido a la generación de grandes volúmenes de datos complejos, fenómeno conocido como big data. El aprovechamiento de esta información permite mejorar la toma de decisiones en salud y, por lo tanto, la calidad de vida de la población. Esta disciplina puede brindarle beneficios muy importantes a Latinoamérica. No obstante, los problemas estructurales, tanto sociales como económicos de la región, sumado a la fragmentación y segmentación de los sistemas de información en salud, representan grandes desafíos para la gestión de los datos. En este artículo se revisarán los aspectos más relevantes de la ciencia de datos en salud en América Latina.*

## ABSTRACT

*Data science is a discipline dedicated to analyzing and processing data with the purpose of turning it into actionable knowledge. In healthcare, data science has had an impact in recent times, mainly due to the generation of large volumes of complex data, a phenomenon known as big data. The use of this information improves decision-making in healthcare and, therefore, quality of life in the general population. This discipline can provide very important benefits to Latin America. However, the region's structural problems, both social and economic, in addition to the fragmentation and segmentation of health information systems, represent great challenges for data management.*

*This article will review the most relevant aspects of health data science in Latin America.*

 Autor para correspondencia

Correo electrónico: [jrosa@hospitalaleman.com](mailto:jrosa@hospitalaleman.com)

<https://doi.org/10.1016/j.rmclc.2022.09.007>

e-ISSN: 2531-0186/ ISSN: 0716-8640/© 2021 Revista Médica Clínica Las Condes.

Este es un artículo Open Access bajo la licencia CC BY-NC-ND  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



## INTRODUCCIÓN

Con el avance de las tecnologías de la información y la comunicación (TICS), se están continuamente generando y almacenando datos en las diferentes esferas sociales. De hecho, esta tendencia ha dado origen al término *big data* (BD) para hacer referencia a aquellos volúmenes de datos demasiado grandes y/o complejos para ser manejados de manera efectiva y/o eficiente por las teorías, tecnologías y herramientas tradicionales relacionadas con los datos<sup>1</sup>. Sin embargo, el tamaño en sí mismo no basta para calificar al BD. Otras características incluyen la velocidad, la heterogeneidad y la variedad de datos<sup>2</sup>.

En el cuidado de la salud, el viraje de los sistemas de información en salud (SIS) desde un enfoque centrado en la enfermedad hacia otro centrado en el paciente ha comenzado a generar grandes volúmenes de información. A las fuentes de datos tradicionales como la historia clínica electrónica (HCE), el registro personal de salud y los estudios complementarios, como imágenes y análisis de laboratorio; se suman otras fuentes no convencionales a partir de la incorporación de nuevas TICS. Estas nuevas tecnologías, prometedoras en relación al aporte de datos, incluyen las ciencias ómicas, la telesalud, los biosensores y los rastreadores de actividad física entre otros. De hecho, se ha empleado el término “informática y analítica en salud” para referirse a las TICS y al análisis de datos necesarios para aprovechar esta información en el cuidado de la salud<sup>3</sup>.

Al momento de procesar y volver comprensible este conjunto enorme y complejo de información es donde la ciencia de datos (CD) comienza a tomar un valor cada vez más importante. A alto nivel y de forma simplificada, podría definirse a la CD como el estudio de los datos<sup>1</sup>. Desde una perspectiva disciplinaria, la CD es un campo interdisciplinario emergente basado en estadística, informática, computación, comunicación, gestión y sociología para estudiar los datos y su entorno. Eventualmente, el análisis de esos datos permitirá transformarlos en conocimientos, decisiones y acciones siguiendo un pensamiento y una metodología determinada<sup>1</sup>.

En este artículo se revisarán los aspectos más relevantes de la CD en salud en Latinoamérica (LA), desde las características de los datos y sus fuentes, hasta los desafíos y oportunidades de esta disciplina en la región.

## CARACTERÍSTICAS DE LOS DATOS EN SALUD

La enorme cantidad y variedad de datos en salud hace que sean interesantes y desafiantes. En cuanto al formato, los datos se pueden clasificar en estructurados y no estructurados. Ejemplos de datos estructurados incluyen datos demográficos (edad, sexo), altura, peso, presión arterial, pruebas de laboratorio y medica-

mentos. Su análisis se puede realizar sin mucho esfuerzo utilizando métodos estadísticos estándar o de aprendizaje automático, ya que los datos ya existen en una estructura fija<sup>4</sup>. Los datos no estructurados, por otro lado, son datos de texto libre (como registros clínicos, resúmenes de alta o informes de estudios), imágenes, audios o videos<sup>5,6</sup>. Se puede extraer mucha información valiosa de los datos, pero es más complicado ya que no están en un formato fácilmente procesable. Por ejemplo, los datos tipo texto libre contienen información contextual como el episodio de atención o la estadía hospitalaria, pero esta información es difícil de extraer. Además, los textos clínicos pueden contener errores gramaticales y ortográficos, ambigüedades y abreviaturas. Este tipo de complejidad aumenta la dificultad de procesar y analizar los datos<sup>4</sup>. A medida que aumenta la cantidad de textos clínicos, los métodos para analizar este tipo de datos, como el procesamiento del lenguaje natural, están ganando un gran interés académico, especialmente entre los científicos de datos.

## INTEROPERABILIDAD DE LOS DATOS EN SALUD

En otro orden de ideas, para que los datos puedan ser compartidos entre diferentes entidades es necesaria la interoperabilidad (IO). En el ámbito de la salud, la IO se define como la capacidad de diferentes SIS para intercambiar datos y usar la información que ha sido intercambiada dentro y a través de los límites de la organización, con el fin de mejorar la prestación efectiva de los cuidados de salud a individuos y comunidades<sup>7</sup>. Según la clasificación propuesta por la *European Telecommunication Standards Institute* (ETSI)<sup>8</sup>, hay cuatro niveles de interoperabilidad:

- **Técnica:** permite la comunicación entre sistemas mediante protocolos e infraestructura necesarios para la comunicación y el intercambio de datos.
- **Sintáctica:** relacionada con la estructuración de la información al momento de intercambiárla. Aquí se diferencian estándares de mensajería como *Health Level Seven* (HL7) o *Digital Imaging and Communication In Medicine* (DICOM) de estándares de documentos como *Continuity of Care Record* (CCR) o *Clinical Document Architecture* (CDA).
- **Semántica:** hace referencia al entendimiento entre sistemas de los códigos intercambiados y el significado de su contenido. Algunos ejemplos de estándares semánticos son *International Classification of Diseases* (ICD), *Systematized Nomenclature of Medicine - Clinical Terms* (SNOMED-CT) o *Logical Observation Identifiers Names and Codes* (LOINC).
- **Organizacional:** es la capacidad de las organizaciones para intercambiar información de manera efectiva.

## FUENTES DE DATOS EN SALUD OFICIALES EN AMÉRICA LATINA

Dentro de la región de LA, una de las fuentes de datos más confiables y transversales a cada región son los datos gubernamentales abiertos. Estos datos son producidos por el gobierno y puestos a disposición de las personas no solo con fines informativos y de lectura, sino también para ser reutilizados en nuevos proyectos, sitios y aplicaciones<sup>9</sup>. Los datos abiertos que proveen los gobiernos deben cumplir con ocho principios básicos, debiendo ser: completos, primarios, actualizados, accesibles, procesables, no discriminatorios, sin propietarios y libres de licencia<sup>9,10</sup>. Varios países de LA ofrecen sus datos en salud en formatos abiertos para qué puedan ser utilizados. Desde el 2011 Brasil cuenta con legislación que aprueba la creación de repositorios digitales de datos abiertos. Argentina y Perú se sumaron en el 2013 y México a partir del 2014. De acuerdo con un reporte de la Organización de las Naciones Unidas para la Educación, la Ciencia y la Cultura (UNESCO), Brasil es el país de LA qué presenta el mayor número de políticas para *data sets* abiertos, seguido por Perú, Argentina, Venezuela, Colombia, México y Bolivia<sup>11</sup>.

Una de las principales fuentes de datos en salud oficiales de LA son las encuestas nacionales. En la mayoría de los países de la región, esta técnica de recolección ya tiene varios años de experiencia, con realizaciones periódicas cada vez más complejas que aportan datos confiables de las diferentes localidades. A través de las mismas, los estados recaban información sobre enfermedades crónicas y sus factores de riesgo, aspectos nutricionales, salud sexual y consumo de sustancias psicoactivas, entre otras. (Tabla 1).

Por otra parte, durante los momentos más álgidos de la pandemia por COVID-19, fueron necesarios datos para comprender el comportamiento de la pandemia, tanto a nivel social como también el perfil clínico de los casos. Para esto, las autoridades locales de distintas regiones reportaron la mayor cantidad de datos posible, cumpliendo con los estándares legales, éticos y de anonimización<sup>28</sup>. Con el paso del tiempo, esas bases de datos fueron evolucionando en cuanto a la cantidad y calidad de sus registros, permitiendo a los gobiernos contar con información confiable para planificar y ejecutar acciones sobre la población a modo de

**Tabla 1. Encuestas nacionales de salud y fuentes de datos relacionadas por país en América Latina**

<b>Argentina</b>	Encuesta Nacional de Factores de Riesgo (ENFR) <sup>12</sup> Encuesta Nacional sobre Salud Sexual y Reproductiva (ENSSyR) <sup>12</sup> Encuesta Nacional sobre Prevalencias de Consumo de Sustancias Psicoactivas (ENPreCoSP) <sup>12</sup>
<b>Bolivia</b>	Encuesta de Demografía y Salud (EDSA) <sup>13</sup>
<b>Brasil</b>	Pesquisa Nacional de Saúde <sup>14</sup>
<b>Chile</b>	Encuesta Nacional de Salud (ENS) <sup>15</sup>
<b>Colombia</b>	Encuesta Nacional de Salud (ENS) <sup>16</sup>
<b>Costa Rica</b>	Encuesta Nacional de Salud (ENSA) <sup>17</sup>
<b>Ecuador</b>	Encuesta Nacional de Salud y Nutrición (ENSANUT) <sup>18</sup>
<b>El Salvador</b>	Encuesta Nacional de Salud (ENS) <sup>19</sup>
<b>Guatemala</b>	Encuesta Nacional de Prevalencia de Enfermedades no Transmisibles y sus Factores de Riesgo Dominio I: Urbano metropolitano, en población de 18 años y más <sup>20</sup>
<b>Honduras</b>	Encuesta Nacional de Demografía y Salud (ENDESA) <sup>21</sup>
<b>México</b>	Encuesta Nacional de Salud y Nutrición (ENSANUT) <sup>22</sup>
<b>Nicaragua</b>	Encuesta de Demografía y Salud (ENDESA) <sup>23</sup>
<b>Panamá</b>	Encuesta Nacional de Salud de Panamá (ENSPA) <sup>24</sup>
<b>Paraguay</b>	Encuesta de salud, bienestar y envejecimiento (SABE) <sup>25</sup>
<b>Perú</b>	Encuesta Demográfica y de Salud Familiar - ENDES <sup>26</sup>
<b>Uruguay</b>	Encuesta Nacional de Salud (ENS) <sup>27</sup>

controlar y prevenir la circulación viral, así como también definir estrategias de acción diferenciadas por región según su nivel de afectación.

Hoy en día, esas fuentes de información se encuentran disponibles para ser consultadas por la población, en presentaciones más fáciles de interpretar rápidamente como informes, reportes y tableros. (Tabla 2).

El proyecto *Salud Urbana en América Latina/Urban Health in Latin America* (SALURBAL)<sup>29</sup>, lanzado en 2017, tiene como objetivo estudiar la salud urbana, la equidad en salud y la sostenibilidad ambiental en las ciudades de LA, para informar las políticas urbanas en todo el mundo. Incluye a todas las ciudades de 100.000 o más habitantes en 2010 en 11 países, para un total de 371 ciudades. Para esto, se obtuvieron registros de mortalidad a nivel individual de cada país durante tantos años como fuese posible. Las causas de muerte se codificaron utilizando los códigos ICD-9 o ICD-10<sup>30</sup>, según el año, y se agruparon en categorías utilizando la clasificación de estimaciones de salud global de la Organización Mundial

de la Salud (OMS). El producto final fue una colección de conjuntos de datos con información sobre cada registro de mortalidad individual, incluido el año, el país, el lugar de residencia, la edad, el sexo, la educación y la causa de muerte. Además, SALURBAL planea compilar encuestas de salud y cualquier estudio de cohorte disponible para desarrollar medidas armonizadas de comportamientos de salud y otros factores de riesgo. Su enfoque inicial ha estado en las encuestas nacionales de salud, principalmente en los factores de riesgo de enfermedades no transmisibles.

En cuanto a iniciativas regionales del uso de datos en salud, Brasil presenta propuestas muy prometedoras. El programa Brasil.IO<sup>31</sup>, es una interfaz de programación de aplicaciones (API) utilizada para recopilar datos de distinta índole. Recibe aportes de voluntarios brasileños que compilan, diariamente, boletines de datos epidemiológicos de las secretarías estatales de salud.

También en Brasil, hay estrategias orientadas a evaluar las políticas de salud pública utilizando análisis que integran herramientas epidemiológicas, estadísticas y de ciencia de datos. Por ejemplo,

**Tabla 2. Fuentes de datos oficiales de COVID-19 por país en América Latina**

<b>Argentina</b>	<a href="https://www.argentina.gob.ar/salud/coronavirus/informacion-epidemiologica">https://www.argentina.gob.ar/salud/coronavirus/informacion-epidemiologica</a>
<b>Bolivia</b>	<a href="https://www.boliviasegura.gob.bo/datos-oficiales">https://www.boliviasegura.gob.bo/datos-oficiales</a>
<b>Brasil</b>	<a href="https://covid.saude.gov.br/">https://covid.saude.gov.br/</a>
<b>Chile</b>	<a href="https://www.gob.cl/coronavirus/cifrasoficiales/">https://www.gob.cl/coronavirus/cifrasoficiales/</a>
<b>Colombia</b>	<a href="https://www.minsalud.gov.co/salud/publica/PET/Paginas/Covid.aspx">https://www.minsalud.gov.co/salud/publica/PET/Paginas/Covid.aspx</a>
<b>Costa Rica</b>	<a href="https://www.ministeriodesalud.go.cr/index.php/vigilancia-de-la-salud/41-lineamientos-coronavirus/527-situacion-nacional-covid-19">https://www.ministeriodesalud.go.cr/index.php/vigilancia-de-la-salud/41-lineamientos-coronavirus/527-situacion-nacional-covid-19</a>
<b>Ecuador</b>	<a href="https://coronavirusecuador.com/">https://coronavirusecuador.com/</a> <a href="https://www.gestionderiesgos.gob.ec/informes-de-situacion-covid-19-desde-el-13-de-marzo-del-2020/">https://www.gestionderiesgos.gob.ec/informes-de-situacion-covid-19-desde-el-13-de-marzo-del-2020/</a>
<b>El Salvador</b>	<a href="https://covid19.gob.sv/">https://covid19.gob.sv/</a>
<b>Guatemala</b>	<a href="https://tablerocovid.mspas.gob.gt/tablerocovid/">https://tablerocovid.mspas.gob.gt/tablerocovid/</a>
<b>Honduras</b>	<a href="http://www.salud.gob.hn/site/">http://www.salud.gob.hn/site/</a>
<b>México</b>	<a href="https://coronavirus.gob.mx/">https://coronavirus.gob.mx/</a>
<b>Nicaragua</b>	<a href="https://www.minsa.gob.ni/index.php/repository/Descargas-MINSA/COVID-19/">https://www.minsa.gob.ni/index.php/repository/Descargas-MINSA/COVID-19/</a>
<b>Panamá</b>	<a href="http://minsa.gob.pa/coronavirus-covid19">http://minsa.gob.pa/coronavirus-covid19</a>
<b>Paraguay</b>	<a href="https://www.mspbs.gov.py/covid-19.php">https://www.mspbs.gov.py/covid-19.php</a>
<b>Perú</b>	<a href="https://covid19.minsa.gob.pe/">https://covid19.minsa.gob.pe/</a>
<b>Uruguay</b>	<a href="https://www.gub.uy/ministerio-salud-publica/coronavirus">https://www.gub.uy/ministerio-salud-publica/coronavirus</a>

mediante el proyecto de evaluación del impacto de las intervenciones de lactancia materna en los hospitales sobre la salud infantil<sup>32</sup>. Este proyecto está liderado por el Instituto de Comunicación e Información Científica y Tecnológica en Salud, junto con el Laboratorio Nacional de Computación Científica (LNCC) y el Centro Federal de Educación Tecnológica de Río de Janeiro como socios; evalúa el impacto en la mortalidad neonatal de los programas de lactancia materna (solo o en combinación)<sup>33</sup>. El proyecto analiza más de 60 millones de partos y alrededor de 320.000 muertes neonatales en maternidades de Brasil durante más de 20 años para comprender mejor la rentabilidad de la adopción combinada de estas políticas públicas que promueven la lactancia materna<sup>34</sup>.

Otro proyecto brasileño centrado en el análisis de datos es sobre la evaluación de la trayectoria de los pacientes a través del sistema de salud mediante el uso de gráficos variables en el tiempo y herramientas de minería de datos<sup>35,36</sup>. Este proyecto, en el que participan investigadores asociados del LNCC, la Universidad Federal de Juiz de Fora y el Ministerio de Salud de Brasil, comprende datos de casi 60 millones de consultas de atención médica o procedimientos de salud realizados a cerca de 6,5 millones de pacientes únicos en la ciudad de San Pablo durante dos años (2014-2015). Los resultados promueven la comprensión de cómo se utiliza el sistema, sus cuellos de botella y dónde emergen, ofreciendo así conocimientos sobre los que construir para mejorar la gestión y asignación de recursos<sup>34</sup>.

Otro país de la región donde se están desarrollando investigaciones mediante el análisis de datos epidemiológicos incorporando algoritmos es México. Un proyecto en curso dirigido por el Instituto de Investigación en Matemática Aplicada y Sistemas de la Universidad Nacional Autónoma de México, que incluye colaboraciones con la Secretaría de Salud de México, ha explorado el uso de modelos epidemiológicos dependientes del tiempo para hacer predicciones confiables sobre la evolución de las pandemias de influenza A (H1N1)<sup>37</sup> y COVID-19. Los resultados obtenidos reflejan el potencial de estas herramientas en la toma de decisiones sanitarias para reducir brotes.

## DESAFÍOS RELACIONADOS A LOS DATOS EN SALUD EN AMÉRICA LATINA

Dentro de LA hay problemas estructurales, tanto económicos como sociales, que afectan a los SIS y a la gestión de los datos en salud. El acceso deficiente a los servicios de atención médica de la población general, los recursos insuficientes y las inequidades sociales impactan a todas las regiones en mayor o menor medida. Paralelamente, la sectorización de la salud (sector público, privado y de seguridad social), la falta de acuerdos regionales y la multiplicidad de procesos asistenciales, generan un entramado complejo que dificulta la gestión de la información.

Por otra parte, un requisito clave para la abstracción y el aprovechamiento de los datos es la interoperabilidad. En LA hay todavía un largo camino por recorrer para alcanzar una interoperabilidad suficiente entre los diferentes SIS locales y regionales. Gran parte de este problema deriva de la falta de TICS necesarias para incorporar estándares técnicos, sintácticos y semánticos; y, peor aún, en la mayoría de las regiones los registros asistenciales continúan siendo en papel, con información fragmentada y duplicada en múltiples archivos físicos. Además, a nivel organizativo, recién hace unos pocos años comenzó a priorizarse la interoperabilidad dentro de la agenda de salud digital, y sólo en algunos países. Argentina, Chile y Uruguay son miembros de la organización SNO-MED-CT<sup>38</sup> y proveen a nivel nacional la licencia gratuita del uso de este estándar de terminología clínica para las instituciones de salud nacionales. Esta estrategia permite codificar la representación del contenido médico, optimizando los datos clínicos y poniéndolos a disposición de diferentes entidades para su análisis posterior.

La mayoría de las bases de datos en salud de acceso público son producidas en países desarrollados. Estos datos representan a una población diferente a la de LA por lo que, extrapolar sus resultados puede conducir a errores. El fomento de la creación de bases de datos locales y regionales facilita que las instituciones de salud puedan acceder y aplicar dicha información a sus procesos, permitiendo que los científicos de datos generen conocimiento de mejor calidad y que, por su parte, las agencias no gubernamentales aborden la situación de salud de la población de forma más integral, proponiendo soluciones a los problemas detectados.

La producción de información en salud está tradicionalmente pensada según el esquema *“dato, información y conocimiento”* (DIC). Este esquema es insuficiente para disponer los datos al servicio de la comunidad, apoyando la toma de decisiones necesarias para disminuir las brechas sociales y entre sectores. En consecuencia, se ha propuesto el esquema *“dato, información, conocimiento, comunicación y acción”* (DICCA)<sup>39</sup>.

Sin embargo, la utilización del BD en el cuidado de la salud, aplicando un esquema DICCA, representa un enorme desafío. Para extraer conocimiento del BD, un SIS requiere herramientas y técnicas de almacenamiento, gestión, análisis y visualización de datos maduras y no convencionales<sup>3</sup>. El volumen y la heterogeneidad de la información, sumado a la necesidad de un procesamiento en tiempo real considerando una gran cantidad de variables, constituyen una meta muy difícil de alcanzar para los SIS en general, y más aun teniendo en cuenta las adversidades que atraviesa la región.

Si bien el camino por recorrer para un aprovechamiento óptimo de los datos en salud representa un gran reto para LA, procurar un trabajo regional colaborativo será fundamental para enfrentar los

problemas que se presenten y progresar en el desarrollo de SIS maduros, capaces de brindar la atención sanitaria que la población se merece.

## CONCLUSIÓN

La CD promete grandes oportunidades en el campo de la salud en LA, fundamentalmente en estos tiempos del BD. A partir de esta disciplina, las entidades de salud podrán generar y gestionar de manera eficiente la información para la toma de decisiones de calidad.

Sin embargo, LA tiene características particulares qué agregan complejidad a la implementación efectiva de la CD en salud. Para conseguir llevarlo a cabo, los gobiernos latinoamericanos deberán enfocarse en desarrollar una visión, misión y plan de acción qué trascienda las fronteras para que el impacto de la disciplina sea transversal en toda la región.

### Declaración de conflicto de interés

Los autores declaran no tener conflictos de interés relacionados a este artículo.

## REFERENCIAS BIBLIOGRÁFICAS

1. Cao L. Data science: A Comprehensive Overview. *ACM Comput Surv.* 2018;1(1):1-42. doi: 10.1145/0000000.00000000
2. Andreu-Perez J, Poon CC, Merrifield RD, Wong ST, Yang GZ. Big data for health. *IEEE J Biomed Health Inform.* 2015;19(4):1193-1208. doi: 10.1109/JBHI.2015.2450362.
3. Pramanik I, Lau RYK, Azad AK, Hossain S, Hossain K, Karmaker BK. Healthcare informatics and analytics in big data. *Expert Syst Appl.* 2020;152:113388. doi: 10.1016/j.eswa.2020.113388
4. Tayefi M, Ngo P, Chomutare T, Dalianis H, Salvi E, Budrionis A, et al. Challenges and opportunities beyond structured data in analysis of electronic health records. *WIREs Comput Stat.* 2021;13:e1549. doi: 10.1002/wics.1549
5. Sun W, Cai Z, Li Y, Liu F, Fang S, Wang G. Data Processing and Text Mining Technologies on Electronic Medical Records: A Review. *J Healthc Eng.* 2018;2018:4302425. doi: 10.1155/2018/4302425.
6. Adnan K, Akbar R. An analytical study of information extraction from unstructured and multidimensional big data. *J Big Data.* 2019;6(1):1-38. doi: 10.1186/s40537-019-0254-8
7. HIMSS. HIMSS Dictionary of Healthcare Information Technology Terms, Acronyms and Organizations, Third Edition. HIMSS; 2013. 294 p.
8. ETSI. European Telecommunication Standards Institute [Internet]. Available from: <https://www.etsi.org>
9. Improving Access to Government through Better Use of the Web [Internet]. [cited 2022 Jul 22]. Available from: <https://www.w3.org/TR/egov-improving/>
10. The 8 Principles of Open Government Data (OpenGovData.org) [Internet]. [cited 2022 Jul 22]. Available from: <https://opengovdata.org/>
11. D'Agostino M, Samuel NO, Sarol MJ, de Cosio FG, Marti M, Luo T, Brooks I, Espinal M. Open data and public health. *Rev Panam Salud Pública.* 2018;42:e66. doi: 10.26633/RPSP.2018.66.
12. INDEC, Instituto Nacional de Estadística y Censos de la República Argentina. [Internet]. [cited 2022 Jul 14]. Available from: <https://www.indec.gob.ar/indec/web/Nivel4-Tema-4-32-68>
13. Encuesta de Demografía y Salud [Internet]. INE. Instituto Nacional de Estadística; 2020 [cited 2022 Jul 21]. Available from: <https://www.ine.gob.bo/index.php/encuesta-de-demografia-y-salud/>
14. IBGE. Instituto Brasileiro de Geografia e Estatística. PNS - Pesquisa Nacional de Saúde. 2019 [Internet]. [cited 2022 Jul 21]. Available from: <https://www.ibge.gov.br/estatisticas/sociais/saude/9160-pesquisa-nacional-de-saude.html?=&t=resultados>
15. MINSAL. Ministerio de Salud de Chile. Departamento de Epidemiología. ENS - Encuesta Nacional de Salud [Internet]. [cited 2022 Jul 14]. Available from: <http://epi.minsal.cl/encuesta-ens/>
16. Ministerio de Salud y Protección Social de Colombia. Páginas - Estudios y encuestas [Internet]. [cited 2022 Jul 21]. Available from: <https://www.minsalud.gov.co/salud/publica/epidemiologia/Paginas/Estudios-y-encuestas.aspx>
17. Hernández K, Saenz I. Primera Encuesta Nacional De Salud (Ensa2006) Informe Técnico Y Primeros Resultados [cited 2022 Jul 21]. Available from: <https://ccp.ucr.ac.cr/farmacoeconomia/documentos/ENSA/informetecnicoypimerosresultadosENSA.pdf>
18. Encuesta Nacional de Salud y Nutrición - ENSANUT - Ministerio de Salud Pública. Ecuador. [cited 2022 Jul 14]. Available from: <https://www.salud.gob.ec/encuesta-nacional-de-salud-y-nutricion-ensanut/>
19. INS. Instituto Nacional de Salud. El Salvador. Encuestas. [cited 2022 Jul 21]. Available from: <http://ins.salud.gob.sv/encuestas/>
20. Departamento de Epidemiología - Home. Gobierno de Guatemala. [cited 2022 Jul 14]. Available from: <https://epidemiologia.mspas.gob.gt/>
21. Encuesta Nacional De Demografía y Salud / Encuesta De Indicadores Múltiples Por Conglomerados (ENDESA/MICS 2019). Available from: <https://www.ine.gob.hn/V3/imag-doc/2021/10/Informe-ENDESA-MICS-2019.pdf>
22. Encuesta Nacional de Salud y Nutrición [Internet]. ENCUESTAS. [cited 2022 Jul 13]. Available from: <https://ensanut.insp.mx/>
23. Instituto Nacional de Información de Desarrollo - INIDE [Internet]. [cited 2022 Jul 21]. Available from: <https://www.inide.gob.ni/Home/endesa>
24. Principal [Internet]. [cited 2022 Jul 21]. Available from: <http://www.gorgas.gob.pa/SIGENSPA/Inicio.htm>
25. Borba A. Encuesta de salud, bienestar y envejecimiento (SABE) [Internet]. Secretaría Técnica de Planificación del Desarrollo Económico y Social. 2018 [cited 2022 Jul 22]. Available from: <https://www.stp.gov.py/v1/sabe/>

26. Instituto Nacional de Estadística e Informática [Internet]. [cited 2022 Jul 14]. Available from: <https://www.gob.pe/institucion/inei/informes-publicaciones/2982736-peru-encuesta-demografica-y-de-salud-familiar-endes-2021>
27. Encuesta Nacional de Salud [Internet]. Ministerio de Salud Pública. [cited 2022 Jul 14]. Available from: <https://www.gub.uy/ministerio-salud-publica/datos-y-estadísticas/datos/encuesta-nacional-salud>
28. Carrillo-Larco RM. COVID-19 data sources in Latin America and the Caribbean. *Travel Med Infect Dis.* 2020;38:101750. doi: 10.1016/j.tmaid.2020.101750.
29. Quistberg DA, Diez Roux AV, Bilal U, Moore K, Ortigoza A, Rodriguez DA, et al.; SALURBAL Group. Building a Data Platform for Cross-Country Urban Health Studies: the SALURBAL Study. *J Urban Health.* 2019;96(2):311-337. doi: 10.1007/s11524-018-00326-0.
30. International Classification of Diseases (ICD) [Internet]. [cited 2022 Jul 22]. Available from: <https://www.who.int/standards/classifications/classification-of-diseases>
31. Home - Brasil.IO [Internet]. [cited 2022 Jul 6]. Available from: <https://brasil.io/home/>
32. Boccolini CS. Assessing the impact of hospital-based breastfeeding interventions on infant health [Internet]. Synapse; 2020. Available from: <https://repo-prod.prod.sagebase.org/repo/v1/doi/locate?id=syn18428446&type=ENTITY>
33. Boccolini CS, Carvalho ML, Oliveira ML. Factors associated with exclusive breastfeeding in the first six months of life in Brazil: a systematic review. *Rev Saude Publica.* 2015;49:91. doi: 10.1590/S0034-8910.2015049005971.
34. Tentori M, Ziviani A, Muchaluat-Saade DC, Favela J. Digital healthcare in Latin America: the case of Brazil and Mexico. *Commun ACM.* 2020;63(11):72-77. doi: 10.1145/3423923
35. Miranda MA, Salvatierra S, Rodríguez I, Álvarez MJ, Rodríguez V. Characterization of the flow of patients in a hospital from complex networks. *Health Care Manag Sci.* 2020;23(1):66-79. doi: 10.1007/s10729-018-9466-2.
36. Rebujo A, Ferreira DR. Business process analysis in healthcare environments: A methodology based on process mining. *Inf Syst.* 2012;37(2):99-116. doi: 10.1016/j.is.2011.01.003
37. Cruz-Pacheco G, Duran L, Esteva L, Minzoni A, Lopez-Cervantes M, Panayotaros P, Ahued Ortega A, Villasenor Ruiz I. Modelling of the influenza A(H1N1)v outbreak in Mexico City, April-May 2009, with control sanitary measures. *Euro Surveill.* 2009;14(26):19254.
38. SNOMED Home page [Internet]. SNOMED. [cited 2022 Jul 22]. Available from: <https://www.snomed.org/#>
39. Alazraqui M, Mota E, Spinelli H. Sistemas de Información en Salud: de sistemas cerrados a la ciudadanía social. Un desafío en la reducción de desigualdades en la gestión local [Health Information Systems: from closed systems to social citizenship. A challenge for the reduction of inequalities in local management]. *Cad. Saude Publica.* 2006;22:2693-2702. doi: 10.1590/s0102-311x2006001200018