

Retos y oportunidades en la accesibilidad de datos

Mónica Calva
mmcalva@utpl.edu.ec
<https://orcid.org/0000-0002-7279-3842>
Hospital UTPL

Recibido: 27/05/2022
Aceptado: 26/09/2022

Resumen

En el área de la salud donde existe una gran variedad y cantidad de información sensible y valiosa no es explotada en su totalidad en beneficio de las personas debido a la falta de acceso a los datos. Hay varios factores que influyen en la accesibilidad de los datos a nivel de organizaciones privadas o públicas, para determinar estos factores se utilizó una revisión bibliográfica, de los 300 artículos consultados, 23 se incluyeron en esta revisión. Entre los factores resultantes que influyen en la accesibilidad de la información se destacaron: gobiernos abiertos, sistemas de información de salud, leyes de protección de datos personales, ética en el uso de la información médica y lagos de datos. Facilitar la accesibilidad a los datos de salud derivaría en una mejora de servicios y planes de tratamientos, ahorro de los estados en recursos económicos, además impulsa la educación y la investigación.

Palabras clave: *accesibilidad, datos, salud, protección, ética.*

Challenges and opportunities in data accessibility

Abstract

In the area of health, where there is a great variety and quantity of sensitive and valuable information, it is not fully exploited for the benefit of people due to the lack of access to data. There are several factors that influence the accessibility of data at the level of private or public organizations, to determine these factors a literature review was used, of the 300 articles consulted, 23 were included in this review. Among the resulting factors that influence the accessibility of information, the following were highlighted: open governments, health information systems, personal data protection laws, ethics in the use of medical information and data lakes. Facilitating accessibility to health data would lead to improved services and treatment plans, savings in economic resources for the states, as well as boosting education and research.

Keywords: *accessibility, data, health, protection, ethics.*

Cómo citar

Calva, M. (2022). Retos y oportunidades en la accesibilidad de datos. *Interfases*, (016), e5888. <https://doi.org/10.26439/interfases2022.n016.5888>

1. Introducción

Hay una gran cantidad de datos que se generan a través de aplicaciones directamente por usuarios y adicionalmente datos generados por organizaciones. Muchos de estos datos no son aprovechados en su totalidad. En el ámbito de la salud hay un sin número de fuentes que generan información de pacientes, como aplicaciones de salud, relojes inteligentes, dispositivos médicos, de monitoreo, signos vitales, equipos de imagen, sistemas registros clínicos electrónicos, etc. Si no están enlazados a un sistema de información no aportan valor o beneficio a las organizaciones. Frecuentemente las técnicas y prácticas para manipulación de datos están enfocadas a datos estructurados, perdiéndose por lo tanto una extensa cantidad de información no estructurada de notas clínicas de texto libre, los cuales pueden ser de gran ayuda para el ámbito de la investigación (Jones et al., 2020) y la educación.

Como consecuencia de manipular estas grandes cantidades de datos, surgen iniciativas de marcos de trabajo sobre la creación de almacenes de datos, lagos de datos. Los lagos de datos es un concepto nuevo que cobra relevancia en los últimos años y que tiene como objetivo ser un depósito centralizado, seleccionado y seguro que almacena todos los datos de las organizaciones, tanto en su forma original como la generada a partir de análisis (Bauer et al., 2021) y con un costo reducido a diferencia de los almacenes de datos.

Sin embargo, surgen preocupaciones del uso ético y legal que puedan darse a los datos recolectados. Deben analizarse las regulaciones legales para la protección de los datos personales y su privacidad. Así como los enfoques éticos usados en proyectos de investigación con este tipo de datos.

Por otro lado, los gobiernos abiertos son iniciativas que han puesto en la mesa de discusión temas de protección de datos, accesibilidad y regulaciones para proteger los derechos de las personas. También abre el campo para discusión sobre la confianza de los titulares de datos sobre el uso que se dé a los mismos. Por lo tanto, los gobiernos son los entes promotores para que se establezcan políticas claras para proteger derechos fundamentales y para generar un valor a los estados a partir de estos datos.

En esta revisión bibliográfica se plantea identificar cuáles son los factores que influyen en la accesibilidad de los datos. En la sección 2 se describen algunos de estos factores y sus implicaciones en el área de la salud. En la sección 3 se presenta sugerencias para investigadores y trabajos futuros y finalmente en la sección 4 las conclusiones.

2. Factores que influyen en la accesibilidad a los datos

Los artículos más relevantes que se han incluido en la revisión bibliográfica se detallan en la Tabla 1.

Tabla 1

Principales artículos de la revisión bibliográfica.

Autores	Artículo	Objetivo / Propósito	Resultados
Alsahafi, Y. A., Gay, V., & Khwaji, A. A.	Factors affecting the acceptance of integrated electronic personal health records in Saudi Arabia: The impact of e-health literacy	Identificar los factores que pueden influir en la aceptación y el uso de un sistema integrado de registros médicos personales en Arabia Saudí.	Bajo nivel de conocimiento entre los ciudadanos saudíes sobre la implantación nacional de un sistema integrado de registros médicos personales y su utilidad.
Avila-Garzon, Cecilia	Applications, methodologies, and technologies for linked open data: A systematic literature review	Realizar una revisión bibliográfica sistemática sobre datos abiertos enlazados (LOD).	La mayoría de los estudios realizados se centran en el uso de tecnologías y herramientas de la web semántica aplicadas a contextos como la biología, las ciencias sociales, las bibliotecas, la investigación y la educación; hay una falta de investigación con respecto a una metodología estandarizada para la gestión de LOD; y se pueden utilizar muchas herramientas para la gestión de LOD, pero la mayoría de ellas carecen de interfaces fáciles de usar para la consulta de conjuntos de datos.
Bauer, D., Froese, F., Garcés-Erice, L., Giblin, C., Labbi, A., Nagy, Z. A., Pardon, N., Rooney, S., Urbanetz, P., Vetsch, P., & Wespi, A.	Building and Operating a Large-Scale Enterprise Data Analytics Platform	Ejecución de una plataforma de procesamiento de datos a gran escala para aplicar la analítica a los datos corporativos a escala en una instancia de nube privada de OpenStack.	Lecciones aprendidas de la construcción y el funcionamiento del datalake empresarial: Automatización, familiaridad de las herramientas, desagregación, uso compartido de recursos, cumplimiento continuo, disponibilidad de datos.

<p>Boudreau LeBlanc, A., Williams-Jones, B., & Aenishaenslin, C.</p>	<p>Bio-Ethics and One Health: A Case Study Approach to Building Reflexive Governance. Frontiers in Public Health</p>	<p>Sentar las bases de un marco metodológico en bioética empírica, cuyos rudimentos se han aplicado a un estudio de caso para construir la gobernanza reflexiva en One Health. Empoderar a los individuos y a las asociaciones en un diálogo con la sociedad, operación que se lleva a cabo mediante un enfoque de estudio de casos sobre sistemas de intercambio de datos.</p>	<p>Un proceso que "reensambla" el paradigma de One Health bajo la perspectiva de la bioética global para crear puentes entre la persona y el ecosistema a través de la ética pragmática.</p>
<p>Buhr, L., Schicktan, S., & Nordmeyer, E</p>	<p>Attitudes Toward Mobile Apps for Pandemic Research Among Smartphone Users in Germany: National Survey</p>	<p>Explorar el potencial de aceptación de las apps orientadas a la investigación (ROA) en la población alemana.</p>	<p>Mayor confianza en las organizaciones gubernamentales (gobierno federal o estatal, oficina regional de salud), públicas (seguro de salud obligatorio) o financiadas por el gobierno (institutos de investigación) para el almacenamiento de datos y la distribución de la aplicación, en contra parte de las organizaciones privadas (instituciones de investigación privadas, clínicas, seguros de salud, empresas de tecnología de la información [TI]). Tener un título universitario aumentaba significativamente la probabilidad de utilizar una aplicación para la pandemia, mientras que tener un origen migratorio la disminuía significativamente. La inmensa mayoría (653/778, 83,9%) de los usuarios de teléfonos inteligentes estaban dispuestos a proporcionar los datos de sus aplicaciones para la investigación financiada por el Estado.</p>
<p>Calva, M., & Piedra, N.</p>	<p>Cycle of Transformation of Open and Linked Medical Data to improve the training of health professionals</p>	<p>Propuesta de un ciclo de transformación de datos médicos a datos abiertos enlazados anónimos</p>	<p>Impulsar mejoras para que los entornos educativos de datos abiertos enlazados contribuyan a mejorar las capacidades de los profesionales de salud. Esto implica mejoras en interfaces de usuario gráficas utilizables que sean fáciles de modificar y de configurar, con datos de pacientes que sirvan para realizar simulaciones clínicas realistas para entrenamiento y evaluación. Apoyo del profesorado, la capacitación y el apoyo financiero. Evaluación de satisfacción del usuario tanto del profesorado como de los estudiantes en estos entornos.</p>

Cohen Glenn, Gerke Sara, & Kramer Daniel B.	Ethical and Legal Implications of Remote Monitoring of Medical Devices	Analizar las implicaciones legales de las normativas de privacidad de EE.UU. y Europa para el acceso de los pacientes a los datos resumidos o brutos de los dispositivos.	Los pacientes estadounidenses pueden tener poco acceso a sus datos brutos recopilados y conservados por los fabricantes de dispositivos en los Estados Unidos en virtud de la Regla de Privacidad de la Ley de Portabilidad y Responsabilidad de los Seguros Médicos, mientras que el Reglamento General de Protección de Datos (RGPD) de la UE concede un mayor acceso a los datos recopilados por los dispositivos cuando el procesamiento de los datos personales entra en el ámbito territorial del RGPD.
Comandè, G., & Schneider, G.	Can the GDPR make data flow for research easier? Yes it can, by differentiating! A careful reading of the GDPR shows how EU data protection law leaves open some significant flexibilities for data protection-sound research activities	Demostrar que la RGPD puede funcionar como facilitador de la investigación.	Los pilares reguladores de la ley europea de protección de datos, el primero relacionado con la protección del derecho fundamental a la protección de datos y el segundo relativo a la promoción de la libre circulación de datos personales, dan lugar a una arquitectura de regímenes de protección de datos por capas, que llegan a reforzar o relajar los derechos de los interesados y las salvaguardias de protección de datos con respecto a las actividades de tratamiento basadas de forma diferente en intereses públicos o meramente económicos. Cada uno de los regímenes de protección de datos identificados da forma a diferentes "puntos normativos habilitantes" para el tratamiento de datos personales sensibles con fines de investigación.
Dunbar, E., Elizabeth Olsen, H., Salomon, E., Bhatt, S., Mutuku, R., Wasunna, B., Edwards, J., Kolko, B., & Holeman, I.	Towards Responsible Data Practices in Digital Health: A case study of an open source community's journey	Plantear una agenda de investigación de diseño que surgió de las necesidades prácticas de una comunidad de código abierto centrada en el software de salud digital para la salud comunitaria en países de ingresos bajos y medios.	Resultados de las etapas formativas de nuestro proceso de diseño en varias fases, que incluyen una revisión bibliográfica de alcance y una reformulación de las políticas y procedimientos institucionales. Una de las principales aportaciones de nuestro estudio de caso es que ofrece un ejemplo del "trabajo previo" institucional necesario para dar sentido al complejo panorama de la protección de datos y trazar un camino para diseñar un software que apoye mejor las prácticas de datos responsables. Este estudio de caso también articula el importante papel de los diseñadores y ejecutores de la sanidad digital en la puesta en práctica de los derechos de los datos de los pacientes.
Eder, J., & Shekhovtsov, V. A.	Data quality for federated medical data lakes	Proponer una arquitectura que ayude a los investigadores a identificar de forma eficiente y eficaz las colecciones de material y datos relevantes con calidad documentada para sus proyectos de investigación, respetando al mismo tiempo las estrictas normas de privacidad.	Este estudio describe el panorama de los biobancos como lagos de datos médicos federados, como las colecciones de muestras y sus anotaciones en la federación europea de biobancos (Biobanking and Biomolecular Resources Research Infrastructure - European Research Infrastructure Consortium, BBMRI- ERIC) y desarrolla un modelo conceptual que captura la información de los esquemas

			<p>con anotaciones de calidad. Este artículo analiza en profundidad las dimensiones de la calidad de los conjuntos de datos para la investigación médica y propone representaciones tanto de los metadatos como de la documentación de la calidad de los datos con el objetivo de ayudar a los investigadores a identificar de forma eficaz y eficiente los conjuntos de datos adecuados para los estudios médicos.</p>
<p>Galeone, C., & Bonzi, R.</p>	<p>OPEN DATA: APPLICATIONS IN SOCIAL CARE AND HEALTH.</p>	<p>Ofrecer una visión general de los datos abiertos y su aplicación en el ámbito social y sanitario. Proporcionamos una definición de datos abiertos, y discutimos el estado actual de la situación en el mundo. Por último, nos centramos en la situación italiana en relación con los datos sanitarios abiertos.</p>	<p>Se ha producido un aumento considerable de la cantidad de datos abiertos, que están cada vez más disponibles debido al elevado número de portales web y conjuntos de datos gratuitos. Sin embargo, no son tantos los conjuntos de datos totalmente abiertos. Existen muchos obstáculos que afectan negativamente a la difusión de los datos abiertos, como por ejemplo: la actualización inadecuada y poco oportuna de los datos; la posibilidad de que la estructura de los datos cambie cuando se publica una actualización; la dificultad de encontrar una codificación única y estándar, con la consiguiente actividad de cotejo costosa y no siempre factible (Bacon y Goldacre, 2020). La escasa disponibilidad de datos completamente abiertos afecta en particular a los datos sanitarios públicos de todo el mundo, sin distinciones relacionadas con el nivel de desarrollo económico y tecnológico del país (World Wide Web Foundation, 2018).</p>
<p>Gefenas, E., Lekstutiene, J., Lukaseviciene, V., Hartlev, M., Mourby, M., & Cathaoir, K.</p>	<p>Controversies between regulations of research ethics and protection of personal data: informed consent at a cross-road.</p>	<p>Explorar las siguientes controversias: ¿Debe considerarse el consentimiento como base jurídica prioritaria para el tratamiento de datos personales en la investigación de datos sanitarios? ¿Puede el consentimiento amplio seguir siendo un fundamento jurídico adecuado para el biobanco? ¿Cuál debería ser el papel de las disposiciones de ética de la investigación que difieren de las normas del RGPD, y cuál debería ser el papel y la función de los comités de ética de la investigación en el entorno cambiante de la investigación de datos sanitarios?</p>	<p>Necesidad de cooperación entre los comités de ética y los organismos de protección de datos. El consentimiento desempeña un papel importante en el mantenimiento de la participación pública en las iniciativas de intercambio de datos. Aunque todavía hay debates en curso sobre cuál sería la mejor opción para el intercambio de datos: el uso del consentimiento dinámico, el consentimiento amplio o la donación de datos sin consentimiento, parece que las preferencias del público por un cierto nivel de control sobre sus datos, independientemente de la base legal en la que se basa el GDPR, apoyan un papel continuo para el consentimiento para mantener la confianza en la investigación, muestra el respeto a los participantes en el estudio, y es una necesidad por defecto bajo las directrices de ética de la investigación.</p>

Healthcare Information and Management Systems Society	2021 HIMSS Healthcare Cybersecurity Survey	Ofrecer una visión del estado de la ciberseguridad en la sanidad basada en las opiniones de 167 profesionales de la ciberseguridad sanitaria. Las organizaciones sanitarias se enfrentan a un sinnúmero de retos, como presupuestos ajustados, infraestructuras envejecidas y un aumento de los ataques de ingeniería social y ransomware.	Los resultados de la Encuesta de Ciberseguridad Sanitaria 2021 de HIMSS sugieren que las organizaciones sanitarias todavía tienen importantes retos que superar. Estos obstáculos al progreso incluyen presupuestos de seguridad ajustados, huellas heredadas cada vez mayores y el creciente volumen de ciberataques y compromisos. Además, los controles de seguridad básicos no se han implementado completamente en muchas organizaciones. Pero quizá la mayor vulnerabilidad sea el factor humano. Las organizaciones sanitarias deberían hacer más para apoyar a los profesionales de la ciberseguridad sanitaria y sus programas de ciberseguridad.
Jones, K. H., Ford, E. M., Lea, N., Griffiths, L. J., Hassan, L., Heys, S., Squires, E., & Nenadic, G.	Toward the development of data governance standards for using clinical free-text data in health research: Position paper.	Elaborar recomendaciones para la creación de normas de gobernanza de datos que se integren en los marcos existentes para el uso de datos personales, a fin de permitir que los datos de texto libre se utilicen de forma segura para la investigación en beneficio de los pacientes y del público.	Guía autorizada sobre la gobernanza de los datos para la reutilización de los datos de texto libre, garantizando la transparencia pública en los flujos y usos de los datos, para tratar los datos de texto libre desidentificados como potencialmente identificables con un uso limitado a los refugios de datos acreditados. Mejora continua para comprender las relaciones entre la eficacia de la desidentificación y los riesgos de reidentificación, y comunicar a todas las partes interesadas.
Kandasamy, K., Srinivas, S., Achuthan, K., & Rangan, V. P.	Digital Healthcare-Cyberattacks in Asian Organizations: An Analysis of Vulnerabilities, Risks, NIST Perspectives, and Recommendations.	Investigar los recientes ciberataques en instituciones sanitarias asiáticas.	Se ha descubierto que hay cinco tipos de ciberataques que dominan las instituciones sanitarias asiáticas. Como parte de este estudio se realiza un análisis detallado de estos ataques, sus vulnerabilidades y los riesgos asociados. Los ciberataques a instituciones sanitarias asiáticas se analizan a través de la lente del marco de riesgos del Instituto Nacional de Normas y Tecnología (NIST). Finalmente, se incluyen algunas recomendaciones únicas para ser utilizadas como medidas de mitigación de riesgos por las organizaciones sanitarias asiáticas y los investigadores para gestionar y mejorar la creciente situación de los ciberataques.
Keefer, P., & Scartascini, C.	Confianza: La clave de la cohesión social y el crecimiento en América Latina y el Caribe	Analizar la confianza a varios niveles sociales y como afecta al crecimiento en América Latina y el Caribe	Diferentes medidas y relaciones de la confianza con economía, civismo, políticas públicas, transformación digital, sociedad, etc. Recomendaciones.

<p>Red Iberoamericana de protección de datos.</p>	<p>Recomendaciones Generales para el Tratamiento de Datos en la Inteligencia Artificial</p>	<p>Recomendaciones para los desarrolladores de productos de IA, con el fin de orientarlos para que desde el diseño del producto se tenga en cuenta las exigencias de las regulaciones sobre tratamiento de datos personales.</p>	<p>Se establecen 10 recomendaciones para desarrolladores de productos de IA</p>
<p>Rooney, S., Bauer, D., Garces-Erice, L., Urbanetz, P., Froese, F., & Tomic, S.</p>	<p>Experiences with managing data ingestion into a corporate datalake.</p>	<p>Describir la gestión de datos, transferencia e ingesta de datos en la plataforma de Datalake.</p>	<p>Explicar la experiencia en el diseño, construcción y gestión de un gran Datalake corporativo.</p>
<p>Shah, S. M., & Khan, R. A.</p>	<p>Secondary use of electronic health record: Opportunities and challenges</p>	<p>Examinar sistemáticamente varios usos secundarios de la historia clínica electrónica con el objetivo de destacar cómo estos usos secundarios afectan a la privacidad de los pacientes. Analizar de forma crítica las normativas GDPR e HIPAA y destacar sus posibles áreas de mejora, teniendo en cuenta el creciente uso de la tecnología y los diferentes usos secundarios de la historia clínica electrónica.</p>	<p>Se describen varios usos secundarios de la historia clínica electrónica (HCE) como la investigación clínica, la vigilancia de la salud pública y las auditorías clínicas para proporcionar servicios de atención sanitaria eficaces, oportunos y de calidad a los pacientes. También se han descrito los desafíos asociados a los usos secundarios de la HCE. Además, se analizan en detalle diferentes cuestiones éticas y de privacidad derivadas de la filtración de datos de HCE. Riesgos para la seguridad de los datos y la privacidad de los pacientes relacionados con los usos secundarios de la HCE, especialmente cuando los datos de la HCE se almacenan en la nube, se transmiten a través de la red y se comparten e intercambian con múltiples partes interesadas. Adicionalmente se estudian importantes normativas de privacidad, como el GDPR y la HIPAA, desde la perspectiva del uso secundario de la HCE. Destacando las posibles áreas de mejora en estas normativas para hacerlas más eficaces en la protección de la privacidad y la seguridad de los datos y para hacerlas robustas frente a las crecientes técnicas asistidas por la IA en el análisis de datos y los ciberataques.</p>

De esta revisión se han obtenido los siguientes aspectos que constituyen un reto u oportunidad para mejorar la accesibilidad a los datos médicos en instituciones públicas o privadas.

2.1 Ley de protección de datos

Un aspecto para tomar en cuenta son las leyes y políticas de protección de datos, que proporcionan estándares para regular el procesamiento de datos que protegen los derechos de las personas y establecen obligaciones para los responsables y custodios de los datos. (Cohen Glenn et al., 2020).

Las leyes de protección de datos recalcan cada vez más que las personas tienen derecho a no estar sujetas a decisiones guiadas únicamente por procesos automatizados. Por todas las implicaciones éticas que esto conlleva.

EE. UU. que ha liderado la promulgación de normativa para protección de datos personales, en 1996 publicó la Ley de Portabilidad y Responsabilidad de los Seguros de Salud (HIPAA), para normar el acceso, y privacidad de los datos de salud de los pacientes. Sin embargo, de acuerdo con Cohen Glenn, la mayoría de los pacientes de EE. UU. pueden tener poco acceso a sus propios datos digitales de salud en los Estados Unidos bajo HIPAA (Cohen Glenn et al., 2020).

Otra regulación vigente desde 2020 en EE. UU. en el estado de California es la Ley de Privacidad del Consumidor de California (CCPA), que propone medidas para proteger mejor la privacidad de los datos de los consumidores. Esta ley tiene un amplio alcance y cubre algunos de los datos que no están cubiertos por HIPAA. Pero CCPA no se aplica a la información de salud protegida que recopilan las entidades cubiertas o sus socios comerciales bajo HIPAA (Cohen Glenn et al., 2020).

Por otro lado, el Reglamento General de Protección de Datos (GDPR) de la Unión Europea (UE); que entró en vigor en el 2018, busca proteger los datos personales y la forma en la que las organizaciones los procesan, almacenan y, finalmente, destruyen, cuando esos datos ya no son requeridos. Algunos países de la región como Ecuador que no disponían de leyes de protección de datos, lograron aprobar en el año 2021 una ley de protección de datos basado en el Reglamento General de Protección de Datos (GDPR) de la Unión Europea.

Si bien la Ley HIPAA proporciona acceso a los registros clínicos del paciente, las leyes que otorgan mayor acceso a los datos recopilados son el Reglamento General de Protección de Datos de la UE y la Ley de Privacidad del Consumidor de California. (Cohen Glenn et al., 2020).

El RGPD proporciona una regulación altamente sofisticada de las actividades de procesamiento de datos con fines de investigación dependiendo de la naturaleza de las actividades de investigación realizadas. Esto fomenta la confianza en los titulares de datos y proporciona un alto nivel de protección de los derechos fundamentales de los interesados (Comandè & Schneider, 2021). También se impulsa para que se cristalicen proyectos de ciencia ciudadana, que permite

incorporar al ciudadano común en proyectos de investigación, estos proyectos deben contar con una normativa específica para la protección de datos personales. (Pierce & Evram, 2022)

En ámbitos como la informática, hay iniciativas de regulaciones diseñadas específicamente para administrar el uso de datos personales, como la Inteligencia Artificial. Por ejemplo, la Red Iberoamericana de Protección de Datos, que consta de 22 autoridades de protección de datos en Portugal y España y en México y otros países de Centro, Sudamérica y el Caribe, ha emitido un reglamento que norma el acceso y uso de los datos en Inteligencia Artificial (IA) a través del reglamento *Recomendaciones Generales para el Tratamiento de Datos Personales en Inteligencia Artificial*. (Red Iberoamericana de protección de datos, 2019) que tiene como objetivo principal evitar amenazas o vulneraciones de los derechos humanos de los titulares de los datos.

2.2 Sistemas de información de salud

Los sistemas de registros médicos electrónicos de las instituciones de salud (EHR), proporcionan beneficios a los pacientes porque son capaces de almacenar y gestionar historiales completos de salud del paciente con información de diferentes áreas como imagen, laboratorio, farmacia, tratamientos clínicos, cirugías y demás procedimientos complementarios; así como información social, financiera y de seguros. Esta centralización de la información permite mejorar la calidad de atención y toma de decisiones ya que reduce los errores en el análisis de datos médicos y toma de decisiones. Facilita la incorporación de herramientas para la toma de decisiones como los sistemas de soporte de decisiones (DSS). Y permite el acceso a los datos desde diferentes ubicaciones y plataformas. (Shah & Khan, 2020). Pero, existen riesgos como la seguridad y privacidad de los datos de los pacientes. De acuerdo con la encuesta de ciberseguridad en el área de salud, realizada por Healthcare Information and Management Systems Society (HIMSS) en el 2021. El 65 % de los encuestados reportaron incidentes de seguridad significativos en sus organizaciones. Ataques de phishing y ransomware fueron los más típicos y frecuentes (Healthcare Information and Management Systems Society, 2021). Los datos médicos están siendo demandados por el cibercrimen y tienen un gran valor en el mercado negro (Kandasamy et al., 2022).

Por otra parte, los registros médicos personales (e-PHR) son cada vez más comunes. Los propios ciudadanos están tomando la iniciativa del uso de plataformas donde se aloja la información de diversos proveedores de salud. Estas iniciativas traen ventajas para los ciudadanos facilitando la centralización de todo su historial de salud en un solo lugar. Permitiéndoles mantener y controlar más fácilmente sus datos clínicos, minimizando errores y duplicidad de información (Alshafi et al., 2022). A pesar de ello, sigue latente los retos para salvaguardar los datos y la privacidad en estos entornos.

Un tema que preocupa son los usos secundarios que se puede dar a la información, en los EHR o e-PHR, como compartir información para la investigación clínica, la vigilancia automatizada

de enfermedades, auditorías clínicas para mejorar la calidad, compartir la información con diferentes agencias gubernamentales, de seguros, etc. Estos usos secundarios también plantean riesgos a la seguridad y privacidad de los datos. Si bien, se incluyen mecanismos como el consentimiento informado o las leyes de protección de datos, las técnicas de vanguardia como el aprendizaje automático (ML), análisis de datos (DA) y piratería hacen que sea aún más difícil proteger completamente la privacidad de un individuo/paciente. (Shah & Khan, 2020). Es indispensable por lo tanto que estos sistemas dispongan de protocolos seguros de almacenamiento, gestión y transmisión de datos.

2.3 Ética en el uso de la información médica

Las técnicas de inteligencia artificial aplicadas en la investigación en el área de la salud, enfrentan cuestiones éticas que deben ser analizadas para evitar una vulneración de la seguridad de los datos y la privacidad de las personas. Desde 1979 se ha establecido principios de respeto, justicia, beneficencia en la investigación con personas, descritos en el Informe de Belmont (The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979) que plantea los principios éticos que deben guiar a investigadores y custodios de datos.

Adicionalmente a las leyes de protección de datos, los comités de ética de la investigación juegan un rol importante para delinear el trabajo de los equipos de investigación. A pesar, de los informes, leyes y reglamentos internos de comités de ética, de acuerdo con (Gefenas et al., 2022) falta una cooperación entre los comités de ética de investigación y los organismos rectores de protección de datos, con el fin de establecer las justificaciones éticas y legales de los estudios de investigación con datos personales. Evitando contradicciones entre normas y proporcionando seguridad y privacidad a los datos personales.

Para lograr este objetivo tiene que crearse normativas a nivel gubernamental. Los gobiernos deben establecer un marco legal y ético para la investigación y el desarrollo. Facilitando mecanismos de comunicación para que los ciudadanos estén informados del beneficio de la investigación basada en datos, y facilite la colaboración a través de asociaciones público-privadas-académicas (Boudreau LeBlanc et al., 2022). Una tarea pendiente en América latina y el Caribe es crear el ambiente para generar confianza en la institucionalidad de los estados, en un reporte publicado por el BID sobre la confianza en la región, los porcentajes en América latina son muy bajos, comparados con países europeos (Keefer & Scartascini, 2022), nueve de cada diez encuestados no confía en los demás. Y cuando se trata de instituciones de gobierno menos de una de cada tres personas confían en las instituciones gubernamentales. En contraparte con países de Europa, por ejemplo, en una encuesta realizada en Alemania del uso de aplicaciones móviles en esta era de pandemia, revela que los ciudadanos confían más en instituciones estatales que en privadas, y un 78,8% de los encuestados están dispuestos a donar datos a instituciones estatales o institutos de investigación para la investigación médica (Buhr

et al., 2022). De acuerdo con el informe la desconfianza genera bajos índices de desarrollo en innovación o investigación. Esto se convierte en una limitante para facilitar el acceso a los datos.

2.4 Repositorios de información de salud - Lagos de datos

Almacenar información de diferentes fuentes es un proceso engorroso y costoso. Con el advenimiento del big data, es un reto conseguir grandes repositorios para guardar información diversa y de diferentes fuentes. Los almacenes de datos hacen estas funciones a través de procesos de ETL (Extract, Transform, Load), lo que implica altos costos en la creación, mantenimiento y gestión de estos entornos. Este es uno de los motivos por el que los lagos de datos se están posicionando como una alternativa para almacenar grandes cantidades de información en su forma y formato original sin costos altos por la gestión de los datos (Eder & Shekhovtsov, 2021).

Se ha demostrado que los lagos de datos permiten a las organizaciones procesar datos a una escala que antes no hubiera sido factible sin sistemas especializados y costosos. Facilitando a las organizaciones almacenar datos sin procesos complejos de ETL, pero sí asociados a un conjunto de políticas, para ser procesado por múltiples equipos diferentes y de acuerdo con sus necesidades (Rooney et al., 2019).

El objetivo principal de un lago de datos es permitir una explotación completa de su contenido para aportar valor a los datos de las organizaciones. Los lagos de datos dentro de un ambiente de un sistema de información deben ser considerados como un nuevo componente del sistema. La diferenciación entre los almacenes de datos y los lagos de datos es que estos últimos son ágiles y flexibles. Con sus propias reglas para el gobierno de datos en términos de gestión de la calidad de los datos, la seguridad de los datos, el ciclo de vida de los datos y los metadatos. Laurent et al. (2020), proponen como un componente clave de la arquitectura del lago de datos el catálogo de metadatos para garantizar la coherencia de las fuentes de datos y su eficiente gestión, evitando así los denominados pantanos de datos.

2.5 Datos abiertos, gobiernos abiertos

Los datos abiertos son fundamentales para los estados e instituciones privadas, academia e institutos de investigación, porque fomentan la transparencia y participación de sus miembros, impulsan el aprendizaje y la investigación. Uno de los primeros gobiernos que introdujeron el concepto de Gobierno Abierto fue el gobierno de Barack Obama (EE.UU.), y ha marcado el punto de inicio para establecer las directrices para la apertura y facilidad de acceso a las actividades gubernamentales y administrativas, con el fin de impulsar acciones prácticas y garantizar un control generalizado sobre la gestión de los bienes públicos por parte de los ciudadanos (Galeone & Bonzi, 2020) .

De acuerdo con la Carta Internacional de Datos Abiertos (Open Data Charter, 2015),

desarrollada en 2015 por algunos gobiernos, sociedad civil y expertos de orden mundial, se establecieron un conjunto de seis principios básicos de datos abiertos:

- 1) *Datos abiertos por defecto*, los gobiernos deben publicar todo tipo de datos excepto aquellos que justificadamente se mantienen cerrados por motivos de seguridad o protección de datos.
- 2) *Oportunos y exhaustivos*, son datos completos, relevantes, publicados en su forma original sin modificaciones.
- 3) *Accesibles y utilizables*, datos fáciles de acceder, en diferentes formatos, gratuitos y legibles por máquinas.
- 4) *Comparables e interoperables*, mayor acceso a conjunto de datos de calidad y estandarización para mejorar la interoperabilidad.
- 5) *Para mejorar la Gobernanza y la Participación Ciudadana*, destaca la importancia de compartir habilidades técnicas y experiencias entre países, y la transparencia en los métodos de recopilación, estandarización y procesos de publicación.
- 6) *Para el Desarrollo Inclusivo y la Innovación*, fomentar la reutilización de datos abiertos para fines comerciales y no comerciales, con el fin impulsar sectores de la educación, salud, medio ambiente, etc.

Todos los datos recopilados pueden terminar en silos de información, sin aportar al objetivo por el que fueron recogidos. Para evitar esto las tecnologías de web semántica proporcionan prácticas y estándares para publicar y conectar los datos posibilitando interoperabilidad de datos, información y sistemas (Calva & Piedra, 2020). Cada vez más se evidencia como los datos vinculados se han ido incorporando en diferentes iniciativas o proyectos que permiten mejorar servicios educativos y de salud para los ciudadanos, investigación médica, intervenciones ambientales y climáticas y un mejor uso de los escasos recursos energéticos, etc. De acuerdo con la revisión bibliográfica de Avila-Garzon (2020) el mayor uso que se ha dado a los datos abiertos enlazados (LOD) es en las áreas de biología y ciencias sociales.

En la revisión bibliográfica realizada por Galeone & Bonzi (2020), mencionan esta evolución significativa sobre la cantidad de datos de salud disponibles gratuitamente para el público en los últimos años. Las investigaciones se centran en el uso de tecnologías de web semántica como el Marco de descripción de recursos (RDF) para que los datos sean accesibles y utilizables. Por ejemplo, la creación de vocabularios RDF: como Observational Health Data Science and Informatics (OHDSI) conectado a recursos vinculados. Conjuntos de datos vinculados sobre el cáncer y recursos biomédicos. Y aplicaciones de la investigación del cáncer El ReVeaLD (un explorador visual en tiempo real y agregador de datos vinculados). En el ámbito farmacéutico Drugle, es un buscador semántico, que lista y analiza información sobre medicamentos disponibles en Internet a través de una simple consulta (Galeone & Bonzi, 2020). Esto puede ser beneficioso para dar un valor agregado a sistemas de registros clínicos o sistemas de aprendizaje de salud y para fortalecer la interoperabilidad entre diferentes organismos.

3. Sugerencias para investigadores y trabajos futuros

Son diversos y complejos los factores que afectan el acceso a la información médica, por ello se sugiere generar grupos de investigación con vinculación social, en donde participen los gobiernos, instituciones privadas y academia. Con el fin de impulsar observatorios de datos o proyectos piloto para crear un ecosistema de datos abiertos médicos.

La academia con sus grupos de investigación debe usar todos los mecanismos vigentes para protección de datos personales, con el fin de salvaguardar la integridad y seguridad de la información personal. Debe existir una estrecha comunicación con los comités de bioética, manejadores de datos y propietarios de datos.

Es indispensable que los investigadores busquen nuevas formas de comunicar y socializar los proyectos, resultados y conclusiones de las investigaciones, para que tenga un mayor alcance de difusión, una ciudadanía bien informada y educada facilita e impulsa la ejecución de proyectos sociales en el ámbito sanitario.

Los trabajos futuros deben ir enfocados a encontrar mejores formas de construir almacenes de datos abiertos colaborativos, a bajo costo y con interfaces fáciles de modificar, configurar y usar por el usuario, para asegurar su uso y aplicabilidad en entornos reales de trabajo. Adicionalmente, los estados deben planificar políticas nacionales estandarizadas que cumplan con estándares de informática médica para la recolección, transformación, procesamiento de información sanitaria, esto contribuirá en gran medida a tener almacenes de datos o lagos de datos abiertos, oportunos y exhaustivos, accesibles y utilizables, comparables e interoperables.

4. Conclusiones

El informe Confianza del BID (Keefer & Scartascini, 2022), muestra como América Latina y el Caribe manifiesta niveles bajos de confianza institucional. Por lo tanto, los gobiernos deben enfocarse en generar políticas en el ámbito de la salud acordes a la realidad de sus sociedades y que generen resultados visibles para el bienestar ciudadano. Con buena política pública, con alianzas estratégicas con la sociedad civil y la academia se puede avanzar en la planificación de proyectos nacionales para datos abiertos en salud. Por otro lado, los gobiernos deben crear un plan de comunicación e información ciudadana para concientizar a la población de las ventajas y oportunidades de los gobiernos abiertos y como la ciudadanía puede contribuir como donante de datos para proyectos de investigación e innovación. Así mismo, se debe asegurar que las leyes, reglamentos y normativas permitan un alto nivel de confianza y privacidad de la información personal. Las leyes como GDPR e HIPPA que adopten los países deben ser ajustadas a las realidades de cada sociedad.

Si bien existe normativa, reglamentos, leyes para protección de datos personales, un aspecto clave es el actuar ético de controladores, manejadores, custodios de datos e investigadores. La

investigación y los enfoques de estructuras tecnológicas deben estar centradas en el ser humano, particularmente en comunidades con niveles más bajos de alfabetización de datos, alfabetización digital o alfabetización en general (Dunbar et al., 2021). Solo así, se puede minimizar los riesgos en la vulneración de los derechos de la persona.

Referencias

- Alsahafi, Y. A., Gay, V., & Khwaji, A. A. (2022). Factors affecting the acceptance of integrated electronic personal health records in Saudi Arabia: The impact of e-health literacy. *Health Information Management Journal*, 51(2), 98–109. <https://doi.org/10.1177/1833358320964899>
- Avila-Garzon, C. (2020). Applications, methodologies, and technologies for linked open data: A systematic literature review. *International Journal on Semantic Web and Information Systems*, 16(3), 53–68. <https://doi.org/10.4018/IJSWIS.2020070104>
- Bauer, D., Froese, F., Garcés-Erice, L., Giblin, C., Labbi, A., Nagy, Z. A., Pardon, N., Rooney, S., Urbanetz, P., Vetsch, P., & Wespi, A. (2021). Building and Operating a Large-Scale Enterprise Data Analytics Platform. *Big Data Research*, 23. <https://doi.org/10.1016/j.bdr.2020.100181>
- Boudreau LeBlanc, A., Williams-Jones, B., & Aenishaenslin, C. (2022). Bio-Ethics and One Health: A Case Study Approach to Building Reflexive Governance. *Frontiers in Public Health*, 10. <https://doi.org/10.3389/fpubh.2022.648593>
- Buhr, L., Schicktanz, S., & Nordmeyer, E. (2022). Attitudes Toward Mobile Apps for Pandemic Research among Smartphone Users in Germany: National Survey. *JMIR MHealth and UHealth*, 10(1). <https://doi.org/10.2196/31857>
- Calva, M., & Piedra, N. (2020). Cycle of Transformation of Open and Linked Medical Data to improve the training of health professionals. *Proceedings of the 15th Latin American Conference on Learning Technologies, LACLO 2020*. <https://doi.org/10.1109/LACLO50806.2020.9381158>
- Cohen, I. G., Gerke, S., & Kramer, D. B. (2020). Ethical and Legal Implications of Remote Monitoring of Medical Devices. *The Milbank Quarterly*, 98(4), 1257–1289. <https://doi.org/10.1111/1468-0009.12481>
- Comandè, G., & Schneider, G. (2021). Can the GDPR make data flow for research easier? Yes it can, by differentiating! A careful reading of the GDPR shows how EU data protection law leaves open some significant flexibilities for data protection-sound research activities. *Computer Law and Security Review*, 41. <https://doi.org/10.1016/j.clsr.2021.105539>
- Dunbar, E., Elizabeth Olsen, H., Salomon, E., Bhatt, S., Mutuku, R., Wasunna, B., Edwards, J.,

- Kolko, B., & Holeman, I. (2021, May 8). Towards Responsible Data Practices in Digital Health: A case study of an open source community's journey. *Conference on Human Factors in Computing Systems - Proceedings*. <https://doi.org/10.1145/3411763.3443438>
- Eder, J., & Shekhovtsov, V. A. (2021). Data quality for federated medical data lakes. *International Journal of Web Information Systems*, 17(5), 407–426. <https://doi.org/10.1108/IJWIS-03-2021-0026>
- Galeone, C., & Bonzi, R. (2020). Open data: Applications in social care and health. *Statistica Applicata - Italian Journal of Applied Statistics*, 32(2), 125–139. <https://doi.org/10.26398/IJAS.0032-008>.
- Gefenas, E., Lekstutiene, J., Lukaseviciene, V., Hartlev, M., Mourby, M., & Cathaoir, K. (2022). Controversies between regulations of research ethics and protection of personal data: informed consent at a cross-road. *Medicine, Health Care and Philosophy*, 25(1), 23–30. <https://doi.org/10.1007/s11019-021-10060-1>
- Healthcare Information and Management Systems Society. (2021). *2021 HIMSS Healthcare Cybersecurity Survey*. <https://www.himss.org/resources/himss-healthcare-cybersecurity-survey>
- Jones, K. H., Ford, E. M., Lea, N., Griffiths, L. J., Hassan, L., Heys, S., Squires, E., & Nenadic, G. (2020). Toward the development of data governance standards for using clinical free-text data in health research: Position paper. *Journal of Medical Internet Research*, 22(6). <https://doi.org/10.2196/16760>
- Kandasamy, K., Srinivas, S., Achuthan, K., & Rangan, V. P. (2022). Digital Healthcare - Cyberattacks in Asian Organizations: An Analysis of Vulnerabilities, Risks, NIST Perspectives, and Recommendations. *IEEE Access*, 10, 12345–12364. <https://doi.org/10.1109/ACCESS.2022.3145372>
- Keefer, P., & Scartascini, C. (2022). *Confianza: La clave de la cohesión social y el crecimiento en América Latina y el Caribe*. <http://dx.doi.org/10.18235/0003911>
- Laurent, A., Laurent, D., & Madera C. (2020). *Data Lakes*. John Wiley & Sons
- Open Data Charter. (2015). *Principios - International Open Data Charter*. <https://opendatacharter.net/principles-es/>
- Pierce, R., & Evram, M. (2022). Getting it right: Implementing data protection in citizen science research. *Insights: The UKSG Journal*, 35. <https://doi.org/10.1629/UKSG.538>
- Red Iberoamericana de protección de datos. (2019). *Recomendaciones Generales para el Tratamiento de Datos en la Inteligencia Artificial*. <https://www.redipd.org/sites/default/files/2020-02/guia-recomendaciones-generales-tratamiento-datos-ia.pdf>
- Rooney, S., Bauer, D., Garces-Erice, L., Urbanetz, P., Froese, F., & Tomic, S. (2019).

Experiences with managing data ingestion into a corporate datalake. *Proceedings - 2019 IEEE 5th International Conference on Collaboration and Internet Computing, CIC 2019*, 101–109. <https://doi.org/10.1109/CIC48465.2019.00021>

Shah, S. M., & Khan, R. A. (2020). Secondary use of electronic health record: Opportunities and challenges. *IEEE Access*, 8, 136947–136965. <https://doi.org/10.1109/ACCESS.2020.3011099>

The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report. Ethical principles and Guidelines for the Protection of Human Subjects of Research*. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>