

RETOS Y OPORTUNIDADES EN LA ACCESIBILIDAD DE DATOS

MÓNICA CALVA
mmcalva@utpl.edu.ec
<https://orcid.org/0000-0002-7279-3842>
Hospital UTPL, Ecuador

RESUMEN. En el área de la salud, existe una gran variedad y cantidad de información sensible y valiosa que no es explotada en su totalidad en beneficio de las personas, debido a la falta de acceso a los datos. Varios factores influyen en la accesibilidad de los datos en organizaciones privadas o públicas, y para determinarlos se ha hecho 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, destacan 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 traería una mejora de servicios y planes de tratamiento, ahorro de recursos económicos estatales, e impulsaría la educación y la investigación.

PALABRAS CLAVE: accesibilidad, datos, salud, protección, ética

CHALLENGES AND OPPORTUNITIES IN DATA ACCESSIBILITY

ABSTRACT. There is a great variety and amount of sensitive and valuable healthcare information that could be used to benefit people, but is not exploited due to limited access to data. This article aims to determine factors influencing data accessibility in private and public organizations through a literature review. Of the 300 related articles identified, 23 were included in this review. Among the resulting factors influencing information accessibility, the following stood out: open governments, health information systems, personal data protection laws, ethics in the use of medical information, and data lakes. Making health data more accessible would improve services and treatment plans, save state resources and promote education and research.

KEYWORDS: accessibility, data, health, protection, ethics

1. INTRODUCCIÓN

Una gran cantidad de datos se genera a través de aplicaciones, ya sea directamente por usuarios o por organizaciones. Muchos de ellos no son aprovechados en su totalidad. En el ámbito de la salud, hay un sinnúmero de fuentes de información de pacientes, por ejemplo, aplicaciones de salud, relojes inteligentes, dispositivos médicos, de monitoreo, signos vitales, equipos de imagen, sistemas de registros clínicos electrónicos, etcétera. Pero 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 en datos estructurados, por lo que se pierde una extensa cantidad de información no estructurada de notas clínicas de texto libre, que puede ser de gran ayuda en el ámbito de la investigación (Jones et al., 2020) y la educación.

Como consecuencia de la necesidad 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 son un concepto nuevo que ha cobrado relevancia en los últimos años, cuyo objetivo es ser un depósito centralizado, seleccionado y seguro que almacena todos los datos de las organizaciones, tanto en su forma original como en la generada a partir de análisis (Bauer et al., 2021), y con un costo reducido, a diferencia de los almacenes de datos.

Sin embargo, existen preocupaciones sobre el uso ético y legal que pueda darse a los datos recolectados. Por ello, 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 sobre la mesa la discusión acerca de la protección de datos, la accesibilidad y las regulaciones para proteger los derechos de las personas. También han comenzado el debate sobre la confianza de los titulares de los datos en el uso que se les dé a estos. Los gobiernos son los entes promotores para que se establezcan políticas claras para proteger los 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 a los datos. En la segunda sección se describen algunos de estos factores y sus implicaciones en el área de la salud. En la tercera sección se presentan sugerencias para investigadores y trabajos futuros; y, finalmente, en la cuarta sección, 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 Saudita.	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, C.	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; pero falta investigación sobre una metodología estandarizada para la gestión de LOD. 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 <i>datalake</i> empresarial: automatización, familiaridad de las herramientas, desagregación, uso compartido de recursos, cumplimiento continuo, disponibilidad de datos.

(continúa)

(continuación)

Autores	Artículo	Objetivo/propósito	Resultados
Boudreau LeBlanc, A., Williams-Jones, B., & Aenishaenslin, C.	Bio-Ethics and One Health: A Case Study Approach to Building Reflexive Governance. <i>Frontiers in Public Health</i>	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.	Un proceso que “reensambla” el paradigma de One Health desde la perspectiva de la bioética global para crear puentes entre la persona y el ecosistema a través de la ética pragmática.
Buhr, L., Schicktanz, S., & Nordmeyer, E.	Attitudes Toward Mobile Apps for Pandemic Research Among Smartphone Users in Germany: National Survey	Explorar el potencial de aceptación de las aplicaciones orientadas a la investigación (ROA) en la población alemana.	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, frente a 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 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.

(continúa)

(continuación)

Autores	Artículo	Objetivo/propósito	Resultados
Calva, M., & Piedra, N.	Cycle of Transformation of Open and Linked Medical Data to Improve the Training of Health Professionals	Propuesta de un ciclo de transformación de datos médicos a datos abiertos enlazados anónimos.	Impulsar mejoras para que los entornos educativos de datos abiertos enlazados contribuyan a optimizar 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 con fines de entrenamiento y evaluación. Apoyo del profesorado, capacitación y soporte financiero. Evaluación de satisfacción del usuario, tanto del profesorado como de los estudiantes, en estos entornos.
Cohen, G., Gerke, S., & Kramer, D. B.	Ethical and Legal Implications of Remote Monitoring of Medical Devices	Analizar las implicaciones legales de las normativas de privacidad de Estados Unidos 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 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 Unión Europea 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	Demostrar que el 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,

(continúa)

(continuación)

Autores	Artículo	Objetivo/propósito	Resultados
	for data protection-sound research activities		que refuerza o relaja los derechos de los interesados y las salvaguardias de protección de datos con respecto a las actividades de tratamiento basadas en intereses públicos o meramente económicos. Cada uno de los regímenes de protección de datos identificados da forma a distintos "puntos normativos habilitantes" para el tratamiento de datos personales sensibles con fines de investigación.
Dunbar, E., Olsen, H. E., 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 <i>software</i> de salud digital para la salud comunitaria en países de ingresos bajos y medios.	Resultados de las etapas formativas de su 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 este 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 <i>software</i> 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	Este estudio describe el panorama de los biobancos como lagos de datos médicos federados, por ejemplo, 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),

(continúa)

(continuación)

Autores	Artículo	Objetivo/propósito	Resultados
		para sus proyectos de investigación, respetando al mismo tiempo las estrictas normas de privacidad.	y desarrolla un modelo conceptual que captura la información de los esquemas 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.
Galeone, C., & Bonzi, R	Open Data: Applications in Social Care and Health	Ofrecer una visión general de los datos abiertos y su aplicación en el ámbito social y sanitario. Proporcionar una definición de datos abiertos y discutir el estado actual de la situación en el mundo. Por último, se centra en la situación italiana en relación con los datos sanitarios abiertos.	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 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 & Goldacre, 2020, como se citó en Galeone & Bonzi, 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, como se citó en Galeone & Bonzi, 2020).

(continúa)

(continuación)

Autores	Artículo	Objetivo/propósito	Resultados
Gefenas, E., Lekstutiene, J., Lukaseviciene, V., Hartlev, M., Mourby, M., & Cathaoir, K.	Controversies between Regulations of Research Ethics and Protection of Personal Data: Informed Consent at a Cross-Road	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?	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 RGPD, apoyan un papel continuo del 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.
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 sinfín de retos, como presupuestos ajustados, infraestructuras	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

(continúa)

(continuación)

Autores	Artículo	Objetivo/propósito	Resultados
		<p>envejecidas y el aumento de los ataques de ingeniería social y <i>ransomware</i>.</p>	<p>sea el factor humano. Las organizaciones sanitarias deberían hacer más para apoyar a los profesionales de la ciberseguridad sanitaria y sus programas de ciberseguridad.</p>
<p>Jones, K. H., Ford, E. M., Lea, N., Griffiths, L. J., Hassan, L., Heys, S., Squires, E., & Nenadic, G.</p>	<p>Toward the Development of Data Governance Standards for Using Clinical Free-Text Data in Health Research: Position Paper</p>	<p>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.</p>	<p>Guía autorizada sobre la gobernanza de los datos para la reutilización de los datos de texto libre, a fin de garantizar 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.</p>
<p>Kandasamy, K., Srinivas, S., Achuthan, K., & Rangan, V. P.</p>	<p>Digital Healthcare -Cyberattacks in Asian Organizations: An Analysis of Vulnerabilities, Risks, NIST Perspectives, and Recommendations</p>	<p>Investigar los recientes ciberataques en instituciones sanitarias asiáticas.</p>	<p>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 del lente del marco de riesgos del Instituto Nacional de Normas y Tecnología (NIST). Finalmente, se incluyen algunas recomendaciones 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.</p>

(continúa)

(continuación)

Autores	Artículo	Objetivo/propósito	Resultados
Keefer, P., & Scartascini, C.	Confianza: la clave de la cohesión social y el crecimiento en América Latina y el Caribe	Analizar la confianza en varios niveles sociales y cómo afecta al crecimiento en América Latina y el Caribe.	Diferentes medidas y relaciones de la confianza con economía, civismo, políticas públicas, transformación digital, sociedad, etcétera. Recomendaciones.
Red Iberoamericana de Protección de Datos	Recomendaciones generales para el tratamiento de datos en la inteligencia artificial	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.	Se establecen diez recomendaciones para desarrolladores de productos de IA.
Rooney, S., Bauer, D., Garcés-Erice, L., Urbanetz, P., Froese, F., & Tomic, S.	Experiences with Managing Data Ingestion into a Corporate Datalake	Describir la gestión de datos, transferencia e ingesta de datos en la plataforma de <i>datalake</i> .	Explicar la experiencia en el diseño, construcción y gestión de un gran <i>datalake</i> corporativo.
Shah, S. M., & Khan, R. A.	Secondary Use of Electronic Health Record: Opportunities and Challenges	Examinar sistemáticamente varios usos secundarios de la historia clínica electrónica con el objetivo de destacar cómo afectan a la privacidad de los pacientes. Analizar de forma crítica las normativas del RGPD y la 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.	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 sus datos se almacenan en la nube, se transmiten a través de la red y se comparten e intercambian con

(continúa)

(continuación)

Autores	Artículo	Objetivo/propósito	Resultados
			múltiples partes interesadas. Adicionalmente, se estudian importantes normativas de privacidad, como el RGPD y la HIPAA, desde la perspectiva del uso secundario de la HCE. Se señalan 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.

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

Las leyes y políticas de protección de datos 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 et al., 2020). Estas leyes 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.

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

Otra regulación vigente desde el 2020 en Estados Unidos 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 considerados por la HIPAA. Pero la CCPA no se aplica a la información de salud protegida que recopilan las entidades cubiertas o sus socios comerciales bajo la HIPAA (Cohen et al., 2020).

Por otro lado, el Reglamento General de Protección de Datos (RGPD) de la Unión Europea, que entró en vigor en el 2018, busca proteger los datos personales y la forma en que las organizaciones los procesan, almacenan y, finalmente, destruyen, cuando 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 basada en el RGPD de la Unión Europea.

Si bien la HIPPA proporciona acceso a los registros clínicos del paciente, las leyes que otorgan mayor acceso a los datos recopilados son el RGPD de la Unión Europea y la CCPA en Estados Unidos (Cohen et al., 2020).

El RGPD proporciona una regulación altamente sofisticada de las actividades de procesamiento de datos con fines de investigación, según la naturaleza de las actividades de investigación realizadas. Esto fomenta la confianza en los titulares de datos y genera un alto nivel de protección de los derechos fundamentales de los interesados (Comandè & Schneider, 2021). También impulsa la puesta en marcha de proyectos de ciencia ciudadana, que permiten incorporar al ciudadano común en proyectos de investigación, los cuales 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 Centroamérica, Sudamérica y el Caribe, ha emitido un reglamento que norma el acceso y uso de los datos en inteligencia artificial (IA) denominado *Recomendaciones generales para el tratamiento de datos personales en inteligencia artificial* (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 de cada paciente con información de diferentes áreas, como imagen, laboratorio, farmacia, tratamientos clínicos, cirugías y demás procedimientos complementarios, además de 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. Igualmente, facilita la incorporación de herramientas para la toma de decisiones, como los sistemas de soporte de decisiones (DSS); y hace posible 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 (HIMSS, 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 porque facilitan la centralización de todo su historial de salud en un solo lugar, lo que les permite mantener y controlar más fácilmente sus datos clínicos, minimizando errores y duplicidad de información (Alsañafi et al., 2022). A pesar de ello, siguen latentes los retos para salvaguardar los datos y la privacidad en estos entornos.

Un aspecto 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, enviar la información a diferentes agencias gubernamentales, de seguros, etcétera. Estos usos secundarios también plantean riesgos para 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 han establecido principios de respeto, justicia y 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 cumplen 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. De esta manera, se evitarían contradicciones entre normas y se proporcionaría seguridad y privacidad a los datos personales.

Para lograr este objetivo, se deben crear normativas en el ámbito gubernamental. Los gobiernos deben establecer un marco legal y ético para la investigación y el desarrollo, así como facilitar mecanismos de comunicación para que los ciudadanos estén informados del beneficio de la investigación basada en datos, de manera que hagan posible 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 Banco Interamericano de Desarrollo (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 confía en las instituciones gubernamentales. En cambio, en países de Europa, por ejemplo, una encuesta realizada en Alemania sobre el 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 cumplen estas funciones a través de procesos de ETL (*extract, transform, load*), que implican altos costos en la creación, mantenimiento y gestión de estos entornos. Este es uno de los motivos por los cuales 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. Facilitan a las organizaciones almacenar datos sin procesos complejos de

ETL, pero sí asociados a un conjunto de políticas, para ser procesados por múltiples equipos y de acuerdo con sus necesidades (Rooney et al., 2019).

El objetivo principal de un lago de datos es hacer posible 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 diferencia entre los almacenes de datos y los lagos de datos es que estos últimos son ágiles y flexibles. Tienen 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 y las 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 de Barack Obama (Estados Unidos), 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 el 2015 por algunos gobiernos, sociedad civil y expertos de orden mundial, se estableció 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.* Son datos fáciles de acceder, en diferentes formatos, gratuitos y legibles por máquinas.
4. *Comparables e interoperables.* Mayor acceso a conjuntos 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, medioambiente, etcétera.

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 la interoperabilidad de datos, información y sistemas (Calva & Piedra, 2020). Cada vez más se evidencia cómo 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, así como un uso más eficiente de los escasos recursos energéticos. 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 su revisión bibliográfica, Galeone y 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 la 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 para salvaguardar la integridad y seguridad de la información personal. Es necesaria 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 tengan 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 estar enfocados en 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 y 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 reporte de confianza del BID (Keefer & Scartascini, 2022) muestra cómo América Latina y el Caribe manifiestan 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 produzcan 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 de cómo la ciudadanía puede contribuir como donante de datos para proyectos de investigación e innovación. Asimismo, se debe asegurar que las leyes, reglamentos y normativas permitan un alto nivel de confianza y privacidad de la información personal. Las leyes como el RGPD o la HIPPA que adopten los países deben ser ajustadas a las realidades de cada sociedad.

Si bien existe normativa, reglamentos y 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., Schick Tanz, 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. En *Proceedings of the 15th Latin American Conference on Learning Technologies (LACLO)* (pp. 1-10). <https://doi.org/10.1109/LACLO50806.2020.9381158>
- Cohen, G., Gerke, S., & Kramer, D. B. (2020). Ethical and legal implications of remote monitoring of medical devices. *The Milbank Quarterly*, 98(4), 1257-1289.
- 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 & Security Review*, 41. <https://doi.org/10.1016/j.clsr.2021.105539>
- Dunbar, E., Olsen, H. E., Salomon, E., Bhatt, S., Mutuku, R., Wasunna, B., Edwards, J., Kolko, B., & Holeman, I. (2021). Towards Responsible Data Practices in Digital Health: A case study of an open source community's journey. En *Extended*

- Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)* (artículo 42, pp. 1-8). Association for Computing Machinery. <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*. Banco Interamericano de Desarrollo. <http://dx.doi.org/10.18235/0003911>
- Laurent, A., Laurent D., & Madera, C. (2020). *Data Lakes*. John Wiley & Sons.
- Open Data Charter. (2015). *Carta Internacional de Datos Abiertos*. <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., Garcés-Erice, L., Urbanetz, P., Froese, F., & Tomic, S. (2019). Experiences with managing data ingestion into a corporate datalake. En *Proceedings-2019 IEEE 5th International Conference on Collaboration and Internet Computing (CIC)* (pp. 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>