Interoperable biomedical repository with FAIR principles for regional collaborative research

Lilian Sosa[1,2]; Lupe Carolina Espinoza[3]; Dely Ramírez[4]
[1] Instituto de Investigaciones en Microbiología (IIM), Facultad de Ciencias, Universidad Nacional Autónoma de Honduras (UNAH), Tegucigalpa, 11101, Honduras. *liliansosa2012@gmail.com*. https://orcid.org/0000-0001-8148-9482

[2] *Grupo de Investigación en Tecnología Farmacéutica, Facultad de Ciencias Químicas y Farmacia,* Universidad Nacional Autónoma de Honduras (UNAH), Tegucigalpa, 11101, Honduras. *liliansosa2012@gmail.com*. https://orcid.org/0000-0001-8148-9482

[3] Departamento de Química, Universidad Técnica Particular de Loja (UTPL), Loja 110107, Ecuador. *lcespinoza@utpl.edu.ec* https://orcid.org/0000-0003-1486-5622

[4] Instituto de Investigaciones en Ciencias Tecnológicas y Aplicadas (IICAT), Universidad Nacional Autónoma de Honduras (UNAH), Tegucigalpa, 11101, Honduras. *Dely.ramirez@unah.edu.hn* . https://orcid.org/0000-0003-1151-1191

Abstract– To propose an interoperable biomedical repository model based on FAIR (Findable, Accessible, Interoperable, Reusable) principles to strengthen collaborative research in Latin America. We conducted a structured narrative review and documentary analysis of regional and global biomedical repositories, peer-reviewed literature published between 2016 and 2025, multilateral body reports, and re3data and OpenDOAR registries. We evaluated semantic interoperability standards, including HL7 FHIR and OMOP CDM. The region presents institutional fragmentation, limited adoption of standardized metadata, and poor semantic interoperability between existing repositories. BIREME's VHL network and PAHO's platforms offer infrastructure that could be articulated into a regional repository. The proposed model comprises four components: institutional ingestion and curation nodes, a federated central catalog, a clinical interoperability module, and distributed governance, together with a five-stage implementation roadmap. Implementing a regional biomedical repository with FAIR principles requires distributed governance, adoption of open standards, and sustainable financing, with potential to reduce research gaps in Central America and Honduras. The model is conceptual and requires empirical validation through a pilot implementation and interoperability testing.

Keywords-- FAIR principles, biomedical interoperability, open data repository, collaborative research, Latin America.

Repositorio biomédico interoperable con principios FAIR para investigación colaborativa regional

Lilian Sosa[1,2]; Lupe Carolina Espinoza[3]; Dely Ramírez[4]

[1] Instituto de Investigaciones en Microbiología (IIM), Facultad de Ciencias, Universidad Nacional Autónoma de Honduras (UNAH), Tegucigalpa, 11101, Honduras. liliansosa2012@gmail.com. https://orcid.org/0000-0001-8148-9482

[2] *Grupo de Investigación en Tecnología Farmacéutica, Facultad de Ciencias Químicas y Farmacia,* Universidad Nacional Autónoma de Honduras (UNAH), Tegucigalpa, 11101, Honduras. liliansosa2012@gmail.com. https://orcid.org/0000-0001-8148-9482

[3] Departamento de Química, Universidad Técnica Particular de Loja (UTPL), Loja 110107, Ecuador. lcespinoza@utpl.edu.ec https://orcid.org/0000-0003-1486-5622

[4] Instituto de Investigaciones en Ciencias Tecnológicas y Aplicadas (IICAT), Universidad Nacional Autónoma de Honduras (UNAH), Tegucigalpa, 11101, Honduras. Dely.ramirez@unah.edu.hn . https://orcid.org/0000-0003-1151-1191

Resumen– ***Proponer un modelo de repositorio biomédico interoperable fundamentado en principios FAIR (Findable, Accessible, Interoperable, Reusable) para fortalecer la investigación colaborativa en América Latina. Se realizó una revisión narrativa estructurada y un análisis documental de repositorios biomédicos regionales y globales, literatura revisada por pares publicada entre 2016 y 2025, informes de organismos multilaterales y registros de re3data y OpenDOAR. Se evaluaron estándares de interoperabilidad semántica, incluidos HL7 FHIR y OMOP CDM. La región presenta fragmentación institucional, adopción limitada de metadatos estandarizados y escasa interoperabilidad semántica entre los repositorios existentes. La red BVS de BIREME y las plataformas de la OPS constituyen infraestructuras articulables para un repositorio regional. El modelo propuesto comprende cuatro componentes: nodos institucionales de ingesta y curaduría, un catálogo central federado, un módulo de interoperabilidad clínica y un esquema de gobernanza distribuida, acompañados de una hoja de ruta de implementación en cinco etapas. La implementación de un repositorio biomédico regional con principios FAIR requiere gobernanza distribuida, la adopción de estándares abiertos y financiamiento sostenible, con el potencial de reducir las brechas de investigación en Centroamérica y Honduras. Se trata de un modelo conceptual que requiere validación empírica mediante una implementación piloto y pruebas de interoperabilidad.***

Palabras clave-- principios FAIR, interoperabilidad biomédica, repositorio de datos abiertos, investigación colaborativa, América Latina.

I. Introducción

La generación de datos biomédicos ha experimentado un crecimiento exponencial durante las últimas dos décadas, impulsado por la expansión de la genómica clínica, la imagenología de alta resolución, los sistemas de registros electrónicos de salud y la proliferación de ensayos clínicos multicéntricos. Este volumen creciente de datos plantea desafíos fundamentales para su almacenamiento, documentación, gestión y reutilización, especialmente en contextos donde la infraestructura tecnológica y el financiamiento científico son limitados [1]. En América Latina, la situación es particularmente compleja: los sistemas de información en salud siguen siendo heterogéneos, la mayoría de los repositorios de datos científicos carecen de estándares comunes y la colaboración interinstitucional se ve obstaculizada por barreras técnicas, administrativas y culturales que restringen el intercambio fluido de información.

Los principios FAIR (Findable, Accessible, Interoperable, Reusable) fueron formulados por Wilkinson y colaboradores en 2016 como un marco internacional de referencia para orientar la gestión y el intercambio de datos científicos [1]. Estos principios establecen que los datos deben ser localizables mediante identificadores persistentes y metadatos ricos, accesibles a través de protocolos estandarizados y abiertos, interoperables mediante vocabularios y ontologías reconocidos por la comunidad científica, y reutilizables con licencias explícitas y documentación suficiente que permita su uso por terceros [2]. Desde su publicación en Scientific Data, los principios FAIR han sido adoptados por organismos financiadores de investigación en Europa y Norteamérica, y han comenzado a permear las políticas de gestión de datos en organismos multilaterales como la UNESCO, la OPS y la Research Data Alliance (RDA) [3], [4].

La implementación efectiva de FAIR en el ámbito biomédico latinoamericano enfrenta obstáculos estructurales que trascienden la adopción de estándares técnicos. La mayoría de las universidades públicas de la región, incluidas las de Centroamérica y Honduras, carecen de repositorios de datos primarios especializados en ciencias de la salud, de esquemas de metadatos estandarizados, de políticas institucionales de datos abiertos y de personal técnico capacitado en gestión de datos de investigación. La producción científica biomédica regional se concentra en publicaciones de acceso abierto, principalmente a través de SciELO y la Biblioteca Virtual en Salud (BVS), y en

repositorios de literatura coordinados por iniciativas como CLACSO [5], [6], [7]; pero el acceso a los datos primarios que respaldan dichas publicaciones sigue siendo muy limitado.

Este panorama genera una brecha epistemológica con impactos concretos en la capacidad científica regional: los estudios no son reproducibles porque los datos no están disponibles públicamente; los metanálisis son difíciles de realizar porque los conjuntos de datos son incompatibles entre sí; y el aprendizaje colectivo derivado de la experiencia clínica acumulada queda confinado a publicaciones parciales que no permiten análisis secundarios rigurosos. En Honduras, donde la investigación biomédica depende en gran medida de la infraestructura de la Universidad Nacional Autónoma de Honduras (UNAH) y de alianzas puntuales con organismos internacionales, esta brecha restringe el aprovechamiento de los datos la vigilancia epidemiológica y los estudios de carga de enfermedad [8].

Frente a este escenario, la propuesta de un repositorio biomédico interoperable con principios FAIR ofrece un camino concreto para la democratización del conocimiento científico en la región. Iniciativas globales como OpenAIRE en Europa, LA Referencia en América Latina y Zenodo del CERN demuestran que es posible construir infraestructuras federadas de datos abiertos, escalables y sostenibles, siempre que existan marcos de gobernanza claros, estándares técnicos compartidos y compromisos institucionales de largo plazo [9]-[11]. La pregunta relevante para la región no es si adoptar estos principios, sino cómo adaptar su implementación a las condiciones tecnológicas, institucionales y financieras específicas de América Latina.

El objetivo de este estudio es elaborar una propuesta metodológica, fundamentada en evidencia, para el diseño de un repositorio biomédico regional interoperable con principios FAIR, identificando los estándares técnicos pertinentes, las infraestructuras existentes susceptibles de articular y los desafíos de gobernanza que deben resolverse para garantizar la sostenibilidad y el uso colectivo del sistema propuesto.

La contribución innovadora del estudio no radica en proponer nuevos principios FAIR ni nuevos estándares de interoperabilidad, sino en su integración y adaptación en una arquitectura regional contextualizada. La propuesta articula infraestructuras latinoamericanas existentes con estándares internacionales de metadatos e interoperabilidad y un esquema de gobernanza distribuida, considerando las limitaciones tecnológicas, institucionales y de sostenibilidad propias de la región. De este modo, el aporte consiste en traducir componentes previamente existentes en un modelo integrado y aplicable como referencia para el desarrollo progresivo de capacidades regionales de gestión e intercambio de datos biomédicos.

II. Materiales y métodos

A. *Tipo de estudio y justificación del diseño*

Este trabajo adopta un diseño de revisión narrativa estructurada con análisis documental y propuesta metodológica basada en evidencia. La elección de este diseño responde a la naturaleza del objeto de estudio: los repositorios biomédicos con principios FAIR en América Latina no constituyen actualmente un fenómeno suficientemente homogéneo ni documentado con datos primarios cuantificables como para permitir un análisis secundario robusto. Sin embargo, existe documentación técnica sustancial sobre repositorios existentes, literatura científica revisada por pares sobre FAIR en biomedicina, e informes institucionales de la OPS, la OMS, la UNESCO y la RDA que permiten construir un análisis documental riguroso y una propuesta metodológica sólida. Los repositorios de datos clínicos con acceso restringido (MIMIC-IV, eICU de PhysioNet) no se incluyeron en la propuesta central dado que requieren Data Use Agreement (DUA) con instituciones norteamericanas, lo que limita su aplicabilidad directa para un modelo latinoamericano autónomo.

B. *Fuentes consultadas*

Las fuentes primarias del análisis incluyeron: (a) artículos científicos publicados en PubMed, PubMed Central y Google Scholar entre 2016 y 2025, en inglés y español; (b) informes técnicos e institucionales de la OPS, la OMS, BIREME, UNESCO y la Research Data Alliance; (c) documentación oficial de estándares de interoperabilidad, incluyendo HL7 FHIR, OMOP CDM, Dublin Core y Data Catalog Vocabulary (DCAT) de W3C; (d) registros de repositorios en re3data.org y OpenDOAR; y (e) portales institucionales de repositorios regionales como LA Referencia, BVS, PAHO IRIS y SciELO. La búsqueda se limitó a fuentes en inglés y español, lo que constituye una limitación de cobertura lingüística.

C. *Criterios de inclusión y exclusión*

Se incluyeron documentos que: (a) abordaran principios FAIR en contextos biomédicos o de ciencias de la salud; (b) describieran repositorios de datos o publicaciones en América Latina y el Caribe; (c) presentaran estándares técnicos de interoperabilidad validados por organismos internacionales; o (d) ofrecieran marcos de gobernanza para ciencia abierta. Se excluyeron documentos anteriores a 2016 (salvo los de carácter fundacional), bases de datos sin documentación técnica verificable, y repositorios comerciales sin política de acceso abierto explicitada. No se incluyeron artículos de opinión sin respaldo empírico.

D. *Estrategia de búsqueda y bases de datos evaluadas*

La búsqueda bibliográfica se realizó con los términos "FAIR data principles biomedical", "open biomedical repository Latin America", "interoperability health data", "HL7 FHIR implementation", "OMOP CDM low-income countries" y "data governance open science". Se combinaron operadores booleanos AND y OR. La búsqueda en re3data.org y OpenDOAR se realizó en abril de 2025, filtrando por país (América Latina), tipo de datos (biomédicos o de salud) y tipo

de acceso (abierto o acceso controlado documentado). Las plataformas y repositorios evaluados se detallan en la Tabla I.

TABLA I
FUENTES, REPOSITORIOS Y ESTÁNDARES CONSIDERADOS EN EL ANÁLISIS DOCUMENTAL

Fuente	Tipo de dato	Dominio	Acceso	Uso	Limitación
BVS / BIREME (OPS/ OMS) [5]	Bibliográfico; literatura científica de salud	Ciencias de la salud	Abierto	Búsqueda bibliográfica y NLP biomédico	Sin arquitectura FAIR para datos primarios; orientado a diseminación bibliográfica
LA Referencia [6]	Repositorio de publicaciones OA; protocolo OAI-PMH	Multi-disciplinario	Abierto	Cosecha de metadatos; infraestructura federada articulable	Sin soporte para datasets primarios; no implementa FAIR pleno
PAHO IRIS (OPS) [12]	Repositorio institucional OPS	Salud pública, epidemiología regional	Abierto	Informes técnicos y guías de la OPS	Alcance limitado a producción institucional OPS
SciELO / SciELO Data [5], [7]	Revistas OA; repositorio datos investigación	Salud y ciencias multidisciplinarias	Abierto	Metanálisis; reproducibilidad científica	SciELO Data en desarrollo regional; baja adopción FAIR en LATAM
re3data. org [13]	Registro de repositorios de datos	Multi-disciplinario	Abierto	Diagnóstico del ecosistema de repositorios regionales	No almacena datos; sólo los registra
Open DOAR [14]	Directorio de repositorios OA	Multi-disciplinario	Abierto	Mapeo de repositorios institucionales de América Latina	No almacena datos; sólo directorio
Zenodo (CERN) [11]	Repositorio generalista de datos de investigación	Multidisciplinario	Abierto / acceso controlado	Depósito de datasets; asignación automática DOI	Generalista; sin soporte FHIR/OMOP nativo
OMOP CDM / OHDSI [15]	Modelo de datos clínicos estandarizado	Datos clínicos, epidemiología observacional	Especificación abierta; datos con acceso controlado (DUA)	Estandarización datos clínicos; estudios multicéntricos	Los datos transformados suelen requerir DUA; e implementación técnica compleja
DataCite Metadata Schema v4.4 [16]	Estándar de metadatos para datos de investigación	Multidisciplinario	Abierto	Descripción, citación y gestión de datasets	Requiere capacitación; adopción limitada en LATAM
WHO IRIS (OMS) [17]	Repositorio institucional OMS	Salud pública global	Abierto	Informes técnicos y guías de la OMS	Alcance limitado a producción institucional OMS

E. Variables, dimensiones y procedimiento de síntesis

Las dimensiones analíticas principales fueron: cumplimiento de principios FAIR (F, A, I, R) según el Modelo de Madurez RDA [3], adopción de estándares de metadatos, uso de identificadores persistentes, protocolos de interoperabilidad, marcos de gobernanza y condiciones de licenciamiento. La síntesis se realizó mediante análisis cualitativo comparativo de las fuentes identificadas, organizando los hallazgos por dimensión y presentando una propuesta de modelo que integra los elementos más robustos identificados en la literatura y la práctica internacional. Véase la Figura 1, mientras que la Figura 2 resume el flujo del análisis documental.

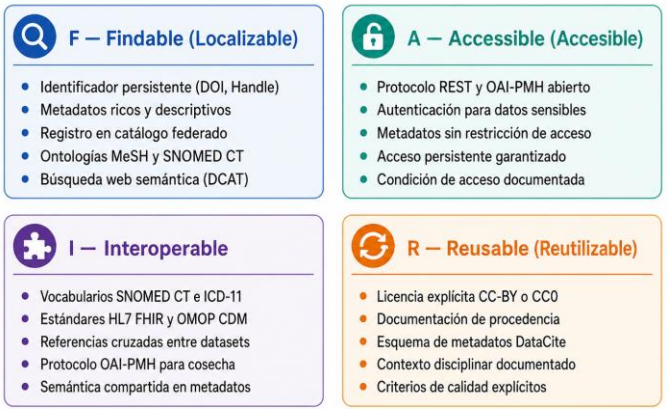


Fig. 1 Principios FAIR aplicados a repositorios biomédicos regionales. Diagrama conceptual de los cuatro principios FAIR con sus criterios específicos en el contexto de repositorios de datos de ciencias de la salud en América Latina.

F. Consideraciones éticas, privacidad y limitaciones metodológicas

Este estudio no involucra datos primarios de personas ni acceso a información clínica identificable. Se trabajó exclusivamente con documentación pública, registros institucionales y literatura científica. No se requirió evaluación por comité de ética. La propuesta fue diseñada conforme a la Recomendación de la UNESCO sobre la Ciencia Abierta [4] y los principios de la RDA para la gestión de datos [3]. El modelo propuesto establece en su componente de módulo clínico el requisito de desidentificación de datos sensibles como condición previa a cualquier depósito, y la necesidad de "Data Use Agreement" para el acceso a datos de nivel individual. Las principales limitaciones metodológicas son: (a) el análisis documental no puede reemplazar un diagnóstico técnico directo de los repositorios existentes; (b) la búsqueda se limitó a fuentes en inglés y español, excluyendo potencialmente documentación en portugués de repositorios brasileños; (c) el estado de los repositorios regionales puede haber cambiado tras la fecha de consulta (abril de 2025); (d) la propuesta de gobernanza es conceptual y no ha sido validada empíricamente; y (e) al tratarse de una revisión narrativa, la selección de fuentes no siguió un protocolo de cribado con doble revisión independiente, por lo que no puede descartarse un sesgo de selección

III. RESULTADOS

A. Panorama de repositorios biomédicos en América Latina

El ecosistema de repositorios de información científica en América Latina está dominado por infraestructuras orientadas a publicaciones (artículos, tesis, informes) y no a datos primarios de investigación. La red LA Referencia articula repositorios institucionales de publicaciones científicas en un conjunto creciente de países latinoamericanos, utilizando el

protocolo OAI-PMH como estándar de interoperabilidad [9], [18]. Aunque LA Referencia constituye una infraestructura federada funcional y de alto valor estratégico, no fue diseñada para almacenar conjuntos de datos biomédicos crudos ni para garantizar la reproducibilidad de los análisis derivados de esos datos.

La Biblioteca Virtual en Salud (BVS), operada por BIREME con el apoyo de la OPS y la OMS, constituye la infraestructura de mayor cobertura regional para información en salud y da acceso a bases de datos bibliográficas (LILACS, MedCarib, IBECS), índices de revistas científicas y recursos educativos en numerosos países de América Latina y el Caribe [5]. Su diseño histórico, orientado a la diseminación bibliográfica, no contempla una arquitectura de gestión de datasets primarios con cumplimiento FAIR.

Las plataformas PAHO IRIS y WHO IRIS almacenan documentos institucionales, informes técnicos, guías y publicaciones de los propios organismos, con metadatos estructurados, pero con un alcance acotado a la producción de la OPS y la OMS respectivamente [17], [12]. SciELO ha dado pasos iniciales hacia la gestión de datos a través de SciELO Data, aunque su alcance en términos de interoperabilidad FAIR está todavía en desarrollo en la región latinoamericana [5], [7]. El análisis de re3data.org y OpenDOAR realizado en abril de 2025 reveló que los repositorios de datos de investigación con foco específico en biomedicina o ciencias de la salud son escasos en América Latina en comparación con Europa y Norteamérica [13], [14]. La búsqueda en re3data.org (abril de 2025) no identificó repositorios nacionales de datos de investigación registrados para Honduras, lo que ilustra la brecha existente incluso respecto a otros países de la región.

B. Evaluación del cumplimiento de principios FAIR en repositorios existentes

El Modelo de Madurez FAIR publicado por la RDA en 2020 proporciona indicadores estandarizados para evaluar el grado de implementación de los principios FAIR en repositorios y conjuntos de datos [3], [19]. Aplicando este modelo como referencia analítica cualitativa, sin asignar puntuaciones formales a sus indicadores, los repositorios latinoamericanos evaluados presentan patrones heterogéneos pero con limitaciones consistentes en varias dimensiones (véase Figura 1):

1) Findable (Localizable): La mayoría de los repositorios bibliográficos regionales asignan metadatos básicos (título, autor, año) pero no implementan identificadores persistentes globales para todos los objetos digitales. La adopción de DOI para artículos es relativamente extendida en SciELO [5], [7], pero es escasa en repositorios institucionales de datos primarios. Los vocabularios controlados y las terminologías biomédicas (MeSH, SNOMED CT, ICD-11) no se integran sistemáticamente en los esquemas de metadatos de los repositorios de la región.

2) Accessible (Accesible): Este principio es el mejor cumplido entre los repositorios analizados. La mayoría permite el acceso gratuito a los metadatos a través de interfaces web y API, siguiendo el principio de que los metadatos deben ser accesibles incluso cuando los datos no lo son [2]. Sin embargo, la disponibilidad de protocolos de autenticación estandarizados y mecanismos de acceso controlado a datos sensibles (datos clínicos identificables) es muy limitada en el contexto regional.

3) Interoperable (Interoperable): Constituye, junto con la reutilización, la brecha más significativa en el ecosistema regional. Los repositorios latinoamericanos usan esquemas de metadatos heterogéneos (Dublin Core básico, formatos propietarios de gestores como DSpace o EPrints) y no implementan estándares de datos biomédicos como HL7 FHIR o OMOP CDM [15], [20]. Esta fragmentación impide la consulta y el análisis federados a escala regional y limita los estudios multicéntricos sin una transformación previa significativa de los datos.

4) Reusable (Reutilizable): Junto con la interoperabilidad, es uno de los principios con menor grado de cumplimiento. La licencia de los datos publicados frecuentemente no se especifica de manera explícita o se omite por completo. La documentación de la procedencia de los datos, los criterios de recolección y las limitaciones de los conjuntos de datos es escasa, lo que dificulta la reutilización con garantías metodológicas suficientes y restringe la posibilidad de realizar análisis secundarios rigurosos por parte de investigadores externos.

C. Estándares de interoperabilidad para un repositorio FAIR regional

La construcción de un repositorio biomédico interoperable requiere la adopción de estándares técnicos validados internacionalmente. Con base en el análisis documental, se identificaron cuatro capas de estándares pertinentes para el contexto latinoamericano:

1) Capa de metadatos: El esquema DataCite proporciona elementos que incluyen identificadores, autores, tipos de datos, fechas, licencias y descripciones, siendo compatible con los principios FAIR y la práctica de repositorios como Zenodo [11], [16]. Dublin Core complementa la descripción básica para repositorios generalistas [21], mientras que el vocabulario DCAT de W3C permite describir catálogos de datos en la web semántica [22].

2) Capa de interoperabilidad clínica: HL7 FHIR es el estándar dominante para el intercambio de datos clínicos en sistemas de salud modernos, con una arquitectura basada en recursos REST que permite integrar datos de registros electrónicos, resultados de laboratorio y ensayos clínicos [20]. El Modelo Común de Datos OMOP, mantenido por OHDSI, permite transformar datos clínicos heterogéneos a un formato estandarizado para estudios epidemiológicos multicéntricos [15].

3) Capa de descubrimiento y cosecha: El protocolo OAI-PMH, ya implementado en LA Referencia, permite la cosecha de metadatos entre repositorios federados de manera sistemática y reproducible [17], [18]. Su integración con APIs REST amplía las posibilidades de consulta automática y

análisis bibliométrico, facilitando la conexión con OpenAIRE y otras infraestructuras globales [10].

4) Capa de identificación y licenciamiento: Los identificadores persistentes DOI (a través de DataCite), ORCID para investigadores y ROR (Research Organization Registry) para instituciones son elementos fundamentales del principio Findable. Las licencias Creative Commons (CC-BY, CC-BY-SA) constituyen el marco jurídico más adoptado internacionalmente para la reutilización de datos científicos abiertos, coherente con la Recomendación sobre Ciencia Abierta de la UNESCO [4].

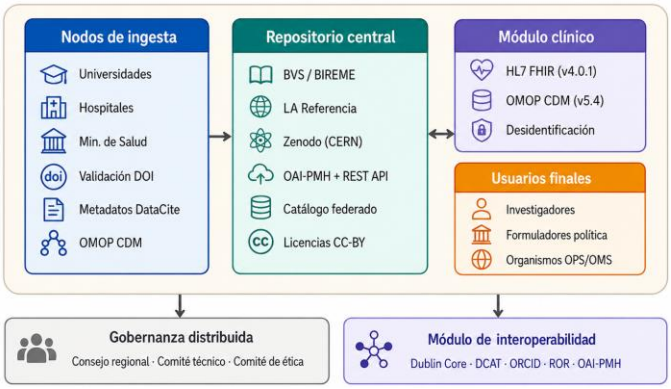


Fig. 2 Flujo metodológico del análisis documental. Secuencia de seis pasos desde la definición del alcance hasta la síntesis de la propuesta de repositorio biomédico FAIR regional.

D. Propuesta de modelo de repositorio biomédico FAIR regional

Con base en los hallazgos anteriores, se propone un modelo de repositorio biomédico interoperable con arquitectura federada que articula infraestructuras existentes y nuevas capacidades bajo principios FAIR. El modelo se organiza en cuatro componentes funcionales (véase Figura 3):

1) Componente 1: Nodos de ingesta y curaduría: Cada institución participante (universidades, hospitales, ministerios de salud) opera un nodo local responsable de recibir, estandarizar y documentar los conjuntos de datos según el esquema DataCite, con herramientas de control de calidad de metadatos, asignación automática de identificadores DOI mediante una membresía consorcial de DataCite y, para datos clínicos, conversión al modelo OMOP CDM.

2) Componente 2: Repositorio central federado: Un nodo central coordina los nodos institucionales mediante OAI-PMH y APIs REST. El catálogo federado permite la búsqueda unificada de datos con filtros por tipo, país, institución, licencia y cobertura temporal. La plataforma Zenodo del CERN ofrece un modelo de referencia viable para este componente [11].

3) Componente 3: Módulo de interoperabilidad clínica: Para datos de naturaleza clínica, el repositorio incorpora una capa de transformación basada en HL7 FHIR y OMOP CDM. Esta capa exige la desidentificación de datos sensibles como condición previa al depósito, la conversión a formatos estandarizados y la generación de documentación de procedencia completa. El acceso a datos de nivel individual requerirá la suscripción de un Data Use Agreement entre las instituciones participantes, en condiciones armonizadas con la normativa de protección de datos de cada país. Cuando dicha normativa impida transferir datos individuales fuera de la institución, el módulo puede operar bajo un esquema de análisis federado, en el que los datos permanecen en cada nodo y solo se comparten resultados agregados, como en la red OHDSI [15].

4) Componente 4: Gobernanza y sostenibilidad: El marco de gobernanza define los roles y responsabilidades de los actores, los mecanismos de resolución de conflictos, los criterios de admisión de nuevos nodos y las políticas de privacidad y financiamiento. Se propone una gobernanza distribuida con consejo directivo regional, comité técnico y comité de ética, inspirada en la experiencia de OpenAIRE en Europa [10].

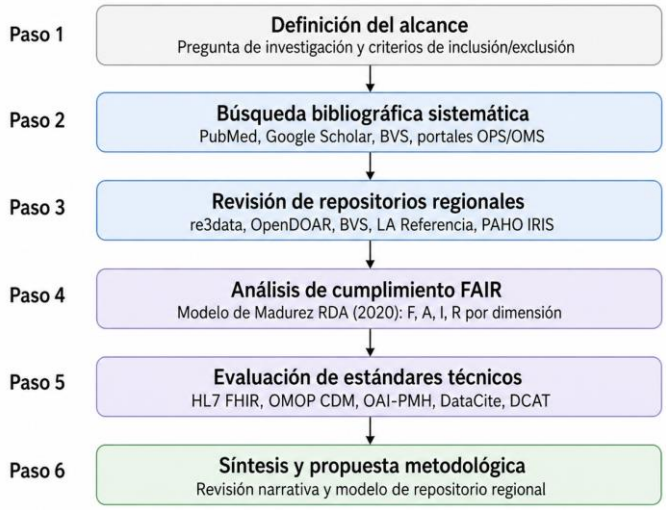


Fig. 3 Arquitectura propuesta del repositorio biomédico interoperable regional. Diagrama estructural del modelo con nodos de ingesta, repositorio central federado, módulo clínico, usuarios finales, gobernanza distribuida e interoperabilidad semántica.

E. Hoja de ruta para la implementación del modelo

A partir de la arquitectura propuesta, se plantea una hoja de ruta de implementación progresiva orientada a trasladar el modelo conceptual hacia una futura validación operativa. La hoja de ruta organiza el proceso en cinco etapas, desde la conformación de la gobernanza y el diagnóstico institucional hasta la evaluación del piloto y el escalamiento regional. Para cada etapa, se identifican los actores responsables, los recursos, los requerimientos tecnológicos básicos y los indicadores de seguimiento (Tabla II).

TABLA II
HOJA DE RUTA PROPUESTA PARA LA IMPLEMENTACIÓN DEL REPOSITORIO BIOMÉDICO FAIR REGIONAL

Etapa	Actividades principales	Responsables	Recursos/requerimientos	Indicadores

1. Gobernanza y planificación	Conformar comité regional; definir políticas de acceso, privacidad, licenciamiento y participación	Universidades, hospitales, ministerios de salud y comités de ética	Personal técnico-jurídico; acuerdos institucionales; políticas de gobernanza	Instituciones participantes; acuerdos formalizados; políticas aprobadas
2. Diagnóstico y preparación técnica	Evaluar infraestructura, datos disponibles, metadatos y capacidades de los nodos	Equipos TI, gestores de datos e instituciones participantes	Inventario tecnológico; personal capacitado; evaluación de calidad de datos	Nodos evaluados; datasets elegibles; brechas técnicas identificadas
3. Implementación piloto	Configurar nodos de ingesta y catálogo federado; implementar DataCite, OAI-PMH/API y mecanismos de identificación	Institución piloto y equipo técnico coordinador	Servidores o infraestructura en nube; software de repositorio; DOI; APIs	Datasets depositados; metadatos completos; disponibilidad del nodo
4. Interoperabilidad y validación	Evaluar intercambio de metadatos y, cuando corresponda, transformación de datos clínicos mediante FHIR/OMOP CDM	Especialistas en informática biomédica, TI y gestores de datos	Herramientas FHIR/OMOP; terminologías; pruebas de interoperabilidad	Intercambios exitosos; cumplimiento FAIR; errores de transformación; calidad de metadatos
5. Evaluación y escalamiento	Evaluar desempeño, aceptación institucional, sostenibilidad y extensión a nuevos nodos	Consejo regional e instituciones participantes	Financiamiento sostenible; capacitación; soporte técnico	Disponibilidad del servicio; instituciones y datasets incorporados; cumplimiento FAIR; sostenibilidad operativa

IV. Discusión

Los resultados de este análisis documental muestran que América Latina dispone de una infraestructura de repositorios orientada principalmente a la diseminación bibliográfica y no a la gestión de datos primarios de investigación bajo principios FAIR. Esta brecha no es exclusiva de la región: a nivel global, la implementación efectiva de FAIR en contextos clínicos enfrenta desafíos técnicos, institucionales y de gobernanza documentados de manera consistente en la literatura especializada [23] - [25]. Lo específico de América Latina es la combinación de fragmentación institucional, limitaciones de financiamiento y heterogeneidad normativa que dificultan la adopción coordinada de estándares comunes.

La comparación con iniciativas europeas, en particular OpenAIRE y el mandato de ciencia abierta de la Comisión Europea que exige planes de gestión de datos como condición de financiación en programas como Horizon Europe [10], [26], ilustra el papel fundamental de las políticas públicas de financiamiento para impulsar la adopción de FAIR. En América Latina, con la excepción parcial de Brasil, Chile y Colombia, esta exigencia no existe de manera generalizada. El RICYT documenta que las inversiones en ciencia y tecnología en Centroamérica siguen siendo marginales en relación con el Producto Interno Bruto, lo que limita estructuralmente la capacidad de construir infraestructura digital sostenible [23].

La propuesta de un repositorio biomédico federado articulado sobre infraestructuras existentes (BVS, LA Referencia, PAHO IRIS) es coherente con las recomendaciones de la UNESCO sobre ciencia abierta, que destacan la importancia de construir sobre capacidades existentes y adaptar los modelos a los contextos locales en lugar de trasplantar soluciones diseñadas para realidades diferentes [4]. FAIRsharing, la iniciativa global para la documentación de estándares y repositorios ofrece un registro curado de estándares, repositorios y políticas que puede orientar su selección progresiva en entornos con recursos limitados [25].

La implementación de HL7 FHIR y OMOP CDM en el módulo de interoperabilidad clínica plantea desafíos específicos para la región. Ambos estándares requieren personal técnico especializado, infraestructura de cómputo adecuada y acuerdos legales claros entre instituciones. La experiencia de la red OHDSI muestra que OMOP CDM puede adoptarse en instituciones de distintos países, pero su implementación requiere inversión inicial en formación y soporte técnico continuo [15]. Para Honduras y el resto de Centroamérica, esto representa un desafío adicional dado el número limitado de profesionales con experiencia en informática biomédica.

Entre las fortalezas de este estudio se cuenta el análisis estructurado de infraestructuras existentes y estándares disponibles, lo que permite construir una propuesta anclada en recursos reales y verificables. Las limitaciones son las inherentes al diseño documental: el análisis no puede reemplazar un diagnóstico técnico directo con acceso a sistemas internos y pruebas de interoperabilidad; la cobertura lingüística excluyó fuentes en portugués; el estado de los repositorios puede haber cambiado tras la consulta (abril de 2025); la propuesta de gobernanza es conceptual, pendiente de negociación entre actores institucionales; y la selección de fuentes, propia de una revisión narrativa, puede estar sujeta a sesgo.. Las líneas futuras de investigación incluyen: (a) un estudio piloto de implementación en al menos un nodo institucional en Honduras o Centroamérica, con el propósito de evaluar empíricamente la factibilidad técnica y operativa del modelo, incluyendo el intercambio de metadatos, la interoperabilidad entre componentes y el cumplimiento de los principios FAIR; (b) la evaluación del cumplimiento FAIR usando los indicadores del modelo RDA en repositorios específicos de la región [19]; y (c) el análisis de marcos legales nacionales de protección de datos y su compatibilidad con principios de ciencia abierta.

En consecuencia, el modelo propuesto debe entenderse como una arquitectura conceptual fundamentada en evidencia documental y no como un modelo validado empíricamente. Su validación futura debería llevarse a cabo por etapas, que incluyan: (i) la evaluación por parte de expertos en informática biomédica, ciencia abierta, gestión de datos y salud pública; (ii) la consulta con posibles partes interesadas institucionales, incluidas universidades, hospitales y autoridades sanitarias; (iii) la implementación piloto en uno o varios nodos institucionales; y (iv) pruebas de interoperabilidad técnica para verificar el intercambio de metadatos y datos utilizando los estándares propuestos. Estas etapas permitirían evaluar la viabilidad técnica, organizativa y operativa del modelo antes de considerar su implementación a nivel regional.

V. Conclusiones

América Latina enfrenta una brecha significativa en la gestión e intercambio de datos biomédicos bajo principios FAIR. Los repositorios existentes en la región, orientados principalmente a publicaciones bibliográficas, carecen de los mecanismos técnicos y los marcos de gobernanza necesarios para garantizar la localización, accesibilidad, interoperabilidad y reutilización de datos primarios de investigación. Esta limitación reduce la capacidad científica regional para realizar análisis secundarios, estudios multicéntricos y revisiones sistemáticas basadas en datos, y coloca a los países de menor capacidad instalada, como Honduras, en una posición de desventaja estructural dentro del ecosistema global de conocimiento biomédico.

El modelo de repositorio biomédico federado propuesto, articulado sobre infraestructuras existentes (BVS, LA Referencia, PAHO IRIS) y estándares internacionales validados (DataCite, HL7 FHIR, OMOP CDM, OAI-PMH), ofrece una alternativa pragmática y escalable para avanzar hacia un ecosistema de datos biomédicos abiertos en la región. Su implementación requiere compromisos institucionales de largo plazo, políticas de financiamiento que incluyan la gestión de datos como componente obligatorio de los proyectos de investigación, y marcos de gobernanza distribuidos que respeten la autonomía de los nodos participantes.

Para la comunidad científica de América Latina, este estudio subraya la urgencia de adoptar principios FAIR no como un fin en sí mismo, sino como una palanca para incrementar la reproducibilidad, la transparencia y el impacto de la investigación biomédica regional. Honduras y Centroamérica podrían beneficiarse de manera particular de un repositorio regional de este tipo: accederían a datos generados por instituciones con mayor capacidad instalada, al tiempo que contribuirían con datos epidemiológicos y clínicos de alto valor para el análisis de enfermedades endémicas, determinantes de salud y carga de enfermedad específica de la región.

Agradecimientos

Agradecemos el apoyo recibido del Instituto de Investigación en Ciencias Médicas y Derecho a la Salud (ICIMEDES) de la Facultad de Ciencias Médicas (FCM) y la Dirección de Investigaciones Científicas, Humanísticas y Tecnológicas (DICIHT) todas ellas de la Universidad Nacional Autónoma de Honduras (UNAH).

Referencias

[1] M. D. Wilkinson et al., "The FAIR Guiding Principles for scientific data management and stewardship", *Sci. Data*, vol. 3, Art. no. 160018, 2016, doi: 10.1038/sdata.2016.18.

[2] GO FAIR Initiative, "FAIR Principles", GO FAIR, 2016. [Online]. Available: https://www.go-fair.org/fair-principles/. Accessed: May 2, 2026.

[3] Research Data Alliance, FAIR Data Maturity Model: Specification and Guidelines, RDA Recommendation, 2020, doi: 10.15497/rda00050.

[4] UNESCO, Recommendation on Open Science. Paris, France: UNESCO, 2021. [Online]. Available: https://unesdoc.unesco.org/ark:/48223/pf0000379949. Accessed: May 2, 2026.

[5] BIREME/PAHO/WHO, "Biblioteca Virtual en Salud", BVS, 2024. [Online]. Available: https://bvsalud.org/. Accessed: May 2, 2026.

[6] CLACSO, "Repositorio Institucional de CLACSO: Red de Bibliotecas Virtuales de Ciencias Sociales en América Latina y el Caribe", Consejo Latinoamericano de Ciencias Sociales, 2024. [Online]. Available: https://biblioteca-repositorio.clacso.edu.ar/. Accessed: May 2, 2026

[7] SciELO, "SciELO Data", SciELO, 2023. [Online]. Available: https://data.scielo.org/. Accessed: May 2, 2026.

[8] Pan American Health Organization, "PLISA: Health Information Platform for the Americas", PAHO/EIH Open Data, 2024. [Online]. Available: https://opendata.paho.org/en. Accessed: May 2, 2026.

[9] LA Referencia, "Red Federada de Repositorios Institucionales de Publicaciones Científicas", LA Referencia, 2024. [Online]. Available: https://www.lareferencia.info/. Accessed: May 2, 2026.

[10] OpenAIRE, "OpenAIRE Guidelines for Data Archives", OpenAIRE Guidelines, 2023. [Online]. Available: https://guidelines.openaire.eu/en/latest/data/index.html. Accessed: May 2, 2026.

[11] Zenodo, "About Zenodo", CERN and OpenAIRE, 2024. [Online]. Available: https://about.zenodo.org/. Accessed: May 2, 2026.

[12] Pan American Health Organization, "PAHO Institutional Repository for Information Sharing (IRIS)", PAHO, 2024. [Online]. Available: https://iris.paho.org/. Accessed: May 2, 2026

[13] re3data.org, "Registry of Research Data Repositories", re3data.org, 2024. [Online]. Available: https://www.re3data.org/. Accessed: May 2, 2026.

[14] OpenDOAR, "Directory of Open Access Repositories", Jisc, 2024. [Online]. Available: https://opendoar.ac.uk/. Accessed: May 2, 2026.

[15] OHDSI, "OMOP Common Data Model v5.4", Observational Health Data Sciences and Informatics, 2024. [Online]. Available: https://ohdsi.github.io/CommonDataModel/. Accessed: May 2, 2026.

[16] DataCite Metadata Working Group, DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs, Version 4.4. DataCite e.V., 2021, doi: 10.14454/3w3z-sa82

[17] World Health Organization, "WHO Institutional Repository for Information Sharing (IRIS)", WHO, 2024. [Online]. Available: https://iris.who.int/. Accessed: May 2, 2026.

[18] C. Lagoze, H. Van de Sompel, M. Nelson, and S. Warner, "The Open Archives Initiative Protocol for Metadata Harvesting", Protocol Version 2.0, Open Archives Initiative, Jun. 14, 2002. [Online]. Available:

https://www.openarchives.org/OAI/openarchivesprotocol.html. Accessed: May 2, 2026.
[19] M. D. Wilkinson et al., "Evaluating FAIR maturity through a scalable, automated, community-governed framework", *Sci. Data*, vol. 6, Art. no. 174, 2019, doi: 10.1038/s41597-019-0184-5.
[20] HL7 International, "FHIR Release 4 (v4.0.1)", HL7 FHIR, 2019. [Online]. Available: https://hl7.org/fhir/R4/. Accessed: May 2, 2026.
[21] Dublin Core Metadata Initiative, "Dublin Core Metadata Element Set, Version 1.1", DCMI, 2012. [Online]. Available: https://www.dublincore.org/specifications/dublin-core/dces/. Accessed: May 2, 2026.
[22] W3C, "Data Catalog Vocabulary (DCAT) - Version 2", W3C Recommendation, Feb. 4, 2020. [Online]. Available: https://www.w3.org/TR/vocab-dcat-2/. Accessed: May 2, 2026.
[23] RICYT, El Estado de la Ciencia 2023: Principales Indicadores de Ciencia y Tecnología Iberoamericanos/Interamericanos. Buenos Aires, Argentina: Red Iberoamericana de Indicadores de Ciencia y Tecnología, 2023. [Online]. Available: https://www.ricyt.org/wp-content/uploads/2023/12/EL-ESTADO-DE-LA-CIENCIA-2023.pdf. Accessed: May 2, 2026.
[24] A. Jacobsen et al., "FAIR Principles: Interpretations and implementation considerations", *Data Intell.*, vol. 2, no. 1–2, pp. 10–29, 2020, doi: 10.1162/dint_r_00024.
[25] S.-A. Sansone et al., "FAIRsharing as a community approach to standards, repositories and policies", Nat. Biotechnol., vol. 37, no. 4, pp. 358–367, Apr. 2019, doi: 10.1038/s41587-019-0080-8.
[26] European Commission, "Open Science", Research and Innovation, European Commission, 2024. [Online]. Available: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/our-digital-future/open-science_en. Accessed: May 2, 2026.