

Los repositorios de datos de investigación: un análisis comparativo

Eder Ávila BarrientosInstituto de Investigaciones Bibliotecológicas y de la Información, Universidad Nacional Autónoma de México (México)  <https://dx.doi.org/10.5209/rgid.94268>

Recibido: 15/2/2024 • Revisado: 15/5/2024 • Aceptado: 20/6/2024

ES Resumen. Los repositorios de datos de investigación son sistemas que facilitan el acceso abierto a los datos de investigación, lo que permite a los investigadores de todo el mundo acceder a ellos en diversos contextos y fomentar la reutilización de los datos de investigación, lo que puede conducir a nuevos descubrimientos científicos y avances tecnológicos; ayudan a preservar los datos de investigación a largo plazo, lo que garantiza que estén disponibles para futuras generaciones de investigadores. En los últimos años, se ha producido un aumento significativo en el número de repositorios de datos de investigación, por lo que resulta trascendental conocer sus atributos y comprender su implementación en el entorno digital. Por lo tanto, el objetivo de este trabajo consiste en analizar las similitudes y diferencias de los repositorios de datos de investigación en función de los criterios establecidos por el movimiento de la ciencia abierta.

Palabras clave. Repositorios de datos de investigación, ciencia abierta, metadatos, datos de investigación, acceso abierto.

ENG Research data repositories: a comparative analysis

ENG Abstract. Research data repositories are systems that facilitate open access to research data, allowing researchers around the world to access it in various contexts; encourage the reuse of research data, which can lead to new scientific discoveries and technological advances; They help preserve research data over the long term, ensuring it is available to future generations of researchers. In recent years, there has been a significant increase in the number of research data repositories, so it is essential to know their attributes and understand their implementation in the digital environment. Therefore, the objective of this work is to analyze the similarities and differences of research data repositories based on the criteria established by the open science movement.

Keywords. Research data repositories, open science, metadata, research data, open access.

Sumario. 1. Introducción. 2. Objetivos y metodología. 3. Revisión de la literatura. 4. Resultados. 5. Discusión. 6. Consideraciones finales. 7. Agradecimientos. 8. Referencias bibliográficas.

Cómo citar: Ávila Barrientos, E. (2024) Los repositorios de datos de investigación: un análisis comparativo, en *Revista General de Información y Documentación* 34 (1), 69-83, e(ID doi). <https://dx.doi.org/10.5209/rgid.94268>

1. Introducción

donde se contempla las publicaciones, los datos, las metodologías, los resultados, los códigos y las herramientas computacionales que se utilizaron en el desarrollo de la investigación por sí misma. De acuerdo con Miedema (2022: 244), en el movimiento de la ciencia abierta tiene presencia el concepto de apertura, esto significa hacer que el proceso científico sea más inclusivo y accesible para todos los actores, dentro y más allá de la comunidad científica.

Los objetivos principales de la ciencia abierta se refieren a cambiar las prácticas científicas para que diseñen su proceso de investigación lo más abierto posible, desde el archivo de datos hasta la revisión por pares abierta (Bautista, Badillo y Lobato, 2023:2).

En este sentido, los datos de investigación son un componente esencial de la investigación científica, y su disponibilidad abierta puede beneficiar a la comunidad científica de diversas maneras. Bajo esta premisa, un análisis comparativo de repositorios de datos de investigación es una herramienta que puede ayudar a los investigadores a identificar los datos más relevantes y útiles para sus necesidades, y a los bibliotecarios a perfilar demandas informativas basadas en el uso de dichos datos. Este tipo de análisis puede realizarse comparando los repositorios de datos en función de una serie de variables, como el área temática, la disciplina científica, el tamaño, el formato, las licencias y las características de los datos.

En este trabajo se abordan los modelos, principios y estándares del enfoque europeo de Ciencia Abierta, debido a que esta región presenta los avances más significativos e integrales en la evolución y reconfiguración del movimiento de Ciencia Abierta en los últimos años.

En años recientes la ciencia abierta ha fomentado la generación de repositorios de datos de investigación en diferentes contextos académicos y científicos, por lo que es necesario comprender cuales son las características de estos sistemas y como contribuyen a la generación de nuevos conocimientos.

2. Objetivos y metodología

2.1. Objetivo general

Este trabajo pretende analizar las similitudes y diferencias de los repositorios de datos de investigación en función de los criterios establecidos por el movimiento de la ciencia abierta.

2.2. Objetivos específicos

Comparar la cobertura de diferentes repositorios de datos de investigación para identificar las políticas de acceso y reutilización de los datos; las herramientas y servicios ofrecidos por los repositorios y la visibilidad y difusión de los datos en dichos repositorios.

2.3. Metodología

La metodología empleada en este trabajo se desarrolló en dos niveles. En el primer nivel, se llevó a cabo un proceso de revisión de la literatura, que llevó a la construcción de una matriz de análisis, con la intención de examinar recursos de información vinculados al objeto de estudio planteado en la investigación. Para ello, se analizaron artículos de investigación, capítulos de libro y estudios de caso.

Estos recursos de información fueron obtenidos en bases de datos, como Elsevier, Springerlink, IEEE y Emerald; en repositorios institucionales, como el repositorio institucional de la UNAM; además, en sitios web académicos e institucionales especializados en el objeto de estudio planteado en la investigación (Research Data Alliance, Research Data Management @Harvard).

La mayor parte de las fuentes utilizadas para el desarrollo de esta investigación son de pago y/o suscripción, esto se debe a la cercana vinculación de los recursos localizados con el objeto de estudio planteado en este trabajo. No obstante, algunas fuentes utilizadas, como es el caso de los repositorios institucionales de acceso abierto, han proporcionado información relevante para el desarrollo de la investigación.

Aunado a ello, la selección de estas fuentes responde a su especificidad temática relacionada con los repositorios de datos de investigación, pues son fuentes que tratan de manera particular el estudio de los repositorios desde diversas aristas, por ejemplo, procedimental, teórica, pragmática y tecnológica. Dentro de estas fuentes, se implementaron estrategias de búsqueda relacionadas con los siguientes términos: repositorios datos de investigación, ciencia abierta, acceso a datos de investigación y sus equivalentes en inglés.

Cabe señalar que, aunque los términos empleados permitieron recuperar recursos que emplearan dichos términos dentro de su título, fue necesario revisar detalladamente cada recurso para identificar el enfoque disciplinario al que estaban dirigidos. Pues la noción conceptual de datos representa a múltiples variables de estudio, desde aquellas que versan en el rubro técnico y pragmático, hasta aquellas que permiten considerar sus nociones epistemológicas.

En el caso de los sitios de *Research Data Alliance* y *Research Data Management @Harvard*, no se realizaron búsquedas específicas de recursos, se llevó a cabo un análisis de la documentación que contenían en sus páginas web que trataran los temas planteados en el objeto de estudio de la investigación.

Por lo tanto, se seleccionaron recursos publicados de 2016 a 2024, tomando en cuenta criterios que reflejan una condicionante tecnológica y cronológica, pues el estudio de los datos se trata de un fenómeno que se encuentra en constante dinamismo debido a la aparición de múltiples publicaciones que analizan su impacto en diferentes contextos y disciplinas. Además, que la mayoría de los recursos utilizados fueron publicados en idioma inglés.

La periodicidad de los años seleccionados responde a la constante actualización de los contenidos y tópicos que se presenta en el objeto de estudio de los datos, pues en años recientes la literatura que versa sobre este tema ha crecido exponencialmente.

Los recursos de información publicados fuera del rango periódico mencionado, fueron elegidos tomando en cuenta su grado de contribución y pertinencia respecto al análisis del objeto de estudio planteado en este trabajo. Es decir, tomando en cuenta su relevancia para obtener una visión global del tema y que se relacionara con los tópicos de organización de la información y el conocimiento.

A su vez, en un segundo nivel, tomando como sustento el método comparativo, se analizó una muestra aleatoria en donde se seleccionaron repositorios de datos de investigación con la finalidad de representarlos en una matriz de análisis, para identificar sus diferencias y similitudes respecto al estudio de su estructura, contenido, recuperación y acceso a los datos de investigación que ofrecen mediante su disponibilidad en el ambiente digital.

Los repositorios de datos de investigación, sujetos a este análisis se seleccionaron tomando en cuenta su relevancia y pertinencia dentro del ecosistema de datos de investigación que se manifiesta en la actualidad, es decir, se contemplaron repositorios que cumplieran con el uso de estándares y directrices internacionales como los principios FAIR y Open Aire.

De esta manera los repositorios de datos de investigación sujetos a este análisis fueron: R3Data, Harvard Dataverse, Zenodo, Mendeley Data, DRYAD, CORA y E-Ciencia.Datos. El análisis metodológico de estos repositorios ha permitido obtener un acercamiento integral a los procesos que intervienen en la representación, organización y sistematización de los datos de investigación, mediante el uso de plataformas digitales.

3. Revisión de la literatura

3.1. Matriz de revisión de literatura

Se llevó a cabo la revisión de literatura relativa al objeto de estudio que se plantea en esta investigación, se identificó la percepción de los repositorios de datos de investigación en el contexto de la ciencia abierta, así como la interacción de los diferentes actores de la Ciencia Abierta (investigadores, instituciones, editores, técnicos) en el marco de los repositorios de datos de investigación.

En la tabla 1, pueden apreciarse las variables que se tomaron en cuenta para el análisis de la literatura efectuado, en donde cabe señalar la integridad entre la percepción de los repositorios de datos de investigación en el dominio de la ciencia abierta y sus respectivos actores. Se destaca la formulación de políticas y directrices que han servido de instrumentos para la toma de decisiones respecto a la aplicación de elementos de la ciencia abierta en el contexto europeo.

La percepción que se tiene acerca de los repositorios de datos de investigación y la ciencia abierta, se encuentra relacionada con una serie de principios tanto conceptuales como instrumentales que refieren a los componentes que forman parte de un contexto integral en donde los investigadores y el desarrollo de políticas de Ciencia Abierta y Acceso Abierto son cruciales para concebir su desarrollo.

Además, los medios de difusión de datos generan diariamente enormes cantidades de datos que son susceptibles de organizarse y sistematizarse a través de los repositorios. Para ello, resulta pertinente el abordaje de metodologías enfocadas al manejo de datos, pues de todo el tsunami de datos disponible en la actualidad, es relevante captar aquellos datos significativos para el desarrollo de nuevos conocimientos o bien, la contrastación de resultados de investigaciones previas que se desarrollan en diversos campos disciplinarios.

Tabla 1 Matriz de revisión de literatura

REFERENCIA	REPORITOSIOS DE DATOS DE INVESTIGACIÓN	CIENCIA ABIERTA	PERCEPCIÓN
ABADAL, E., et al., 2023. Ciencia abierta en España 2023: informe de situación y análisis de la percepción. Barcelona-València: Grupo Ciencia Abierta. http://hdl.handle.net/2445/200020		La percepción de que la ciencia abierta puede dificultar la evaluación de la investigación. La creciente presión de la sociedad para que la investigación sea más abierta y transparente. Desarrollar políticas institucionales que promuevan la ciencia abierta.	Los investigadores españoles están a favor de la ciencia abierta, pero perciben una serie de barreras que dificultan su implementación: La falta de financiación para la publicación en acceso abierto. La falta de políticas institucionales que promuevan la ciencia abierta.
PAIC, A. (2021), Open Science - Enabling Discovery in the Digital Age, Going Digital Toolkit Note, No. 13, https://goingdigital.oecd.org/data/notes/No13_ToolkitNote_OpenScience.pdf .	La infraestructura y las herramientas científicas abiertas, como los repositorios institucionales, los repositorios de datos y el software de código abierto, facilitan la difusión y utilización del conocimiento científico.	La ciencia abierta fomenta la colaboración interdisciplinaria, lo que permite a investigadores de diferentes campos compartir experiencia, conocimientos y datos, lo que conduce a nuevos descubrimientos y avances.	Los investigadores necesitan capacitación y apoyo para utilizar eficazmente herramientas y prácticas de ciencia abierta, cerrar la brecha digital y garantizar el acceso equitativo a los recursos de ciencia abierta.

European Commission, Directorate-General for Research and Innovation, Open innovation, open science, open to the world – A vision for Europe, Publications Office, 2015, https://data.europa.eu/doi/10.2777/061652	Los datos y protocolos abiertos promueven la reproducibilidad, asegurando la confiabilidad y validez de los hallazgos científicos.	La ciencia abierta promueve el acceso abierto, el intercambio y la reutilización de los resultados de la investigación científica, incluidas publicaciones, datos, software y protocolos. Este enfoque mejora la colaboración, acelera el descubrimiento, fortalece la reproducibilidad y amplía el acceso público al conocimiento científico.	Establecer planes claros de gestión de datos, capacitar a los investigadores en conservación de datos y desarrollar políticas de propiedad e intercambio de datos. Brindar capacitación y apoyo a investigadores, bibliotecarios, formuladores de políticas y ciudadanos sobre innovación abierta, ciencia abierta y colaboración global.
Open Research Data: From Vision to Practice SpringerLink, [en línea]. Recuperado a partir de : https://link.springer-com.pbidi.unam.mx:2443/chapter/10.1007/978-3-319-00026-8_14 [accedido 8 noviembre 2023].	Los datos abiertos fomentan la colaboración interdisciplinaria, lo que permite a los investigadores de diferentes campos compartir experiencia, conocimientos e información, lo que conduce a nuevos descubrimientos y avances.		Los investigadores necesitan formación y apoyo para utilizar eficazmente las herramientas y prácticas de datos abiertos, lo que ayuda a superar la brecha digital y garantiza el acceso equitativo a los recursos de datos abiertos.
Science, Digital; Hahnel, Mark; Smith, Graham; schoenenberger, henning; Scaplehorn, Niki; Day, Laura (2023). The State of Open Data 2023. Digital Science. Report . https://doi.org/10.6084/m9.figshare.24428194.v1	Las tendencias apuntan a un futuro más abierto y colaborativo para la investigación, tomando en cuenta: El aumento de la adopción de políticas de acceso abierto por parte de los gobiernos y las instituciones de investigación. El aumento de la disponibilidad de datos abiertos. El desarrollo de nuevas herramientas y tecnologías para facilitar la ciencia abierta.	La falta de financiación para la publicación en acceso abierto es una de las principales barreras para la ciencia abierta. La falta de políticas institucionales que promuevan la ciencia abierta es otra barrera importante. La falta de formación y concienciación sobre la ciencia abierta también es un desafío.	El 77% de los investigadores cree que la ciencia abierta es importante o muy importante. El 68% de los investigadores cree que la ciencia abierta está mejorando la calidad y la transparencia de la investigación. El 63% de los investigadores cree que la ciencia abierta está facilitando la colaboración entre investigadores.

Tabla de elaboración propia, 2023.

3.2. Ciencia abierta

La percepción que se tiene acerca de los repositorios de datos de investigación y la ciencia abierta, se encuentra relacionada con una serie de principios tanto conceptuales como instrumentales que refieren a los componentes que forman parte de un contexto integral en donde los investigadores y el desarrollo de políticas de Ciencia Abierta y Acceso Abierto son cruciales para concebir su desarrollo.

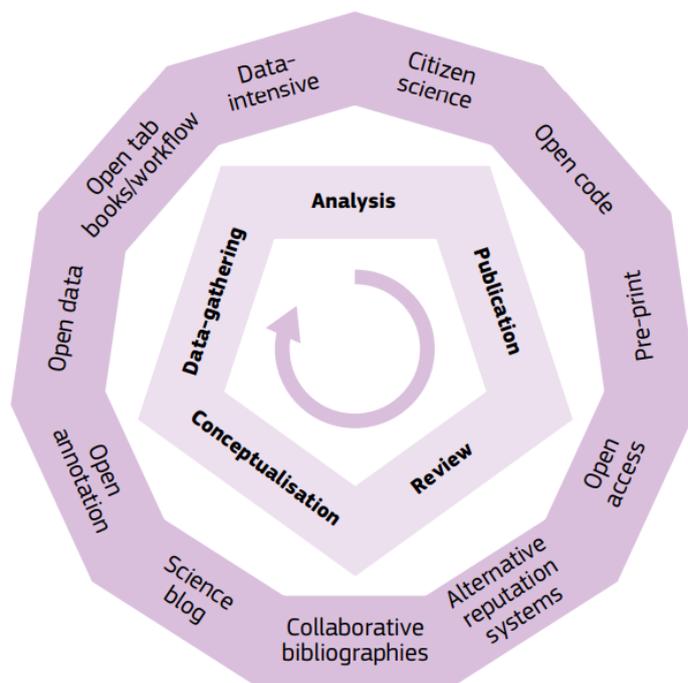
Además, los medios de difusión de datos generan diariamente enormes cantidades de datos que son susceptibles de organizarse y sistematizarse a través de los repositorios. Para ello, resulta pertinente el abordaje de metodologías enfocadas al manejo de datos, pues de todo el tsunami de datos disponible en la actualidad, es relevante captar aquellos datos significativos para el desarrollo de nuevos conocimientos o bien, la contrastación de resultados de investigaciones previas que se desarrollan en diversos campos disciplinarios.

En la figura 1 puede apreciarse que este impacto está relacionado desde el inicio de la investigación hasta la publicación de los resultados, y sobre cómo se organiza este ciclo. El círculo exterior en la Figura 1 muestra la nueva naturaleza interconectada de la Ciencia Abierta, mientras que el círculo interior muestra todo el proceso científico, desde la conceptualización de ideas de investigación hasta su publicación.

Cada paso en el proceso científico está vinculado a los cambios continuos provocados por la Ciencia Abierta, como la aparición de sistemas alternativos para establecer reputación científica, cambios en la manera de evaluar la calidad y el impacto de la investigación, así como el uso creciente de blogs científicos y acceso abierto a datos y publicaciones de diversa naturaleza.

De esta manera, la UNESCO (2021: 7), en la *Recommendation of Open Science*, destaca el carácter inclusivo de la Ciencia Abierta al considerarla como un concepto que combina diversos movimientos y prácticas destinados a hacer que el conocimiento científico multilingüe esté disponible de forma abierta, sea accesible y reutilizable para todos, para aumentar las colaboraciones científicas y el intercambio de información en beneficio de la ciencia y la sociedad, y abrir los procesos de creación, evaluación y comunicación del conocimiento científico a los agentes sociales más allá de la comunidad científica tradicional.

Por lo que la visión holística de la Ciencia Abierta permite considerar elementos que integran su conceptualización en un entorno complejo, como lo es el de la información científica y académica que se genera actualmente, pues la mayoría de esta información fundamenta su acción en el desarrollo de políticas.



(fig.1) Ciencia abierta: abriendo el proceso de investigación. Fuente: Dirección General de Investigación e Innovación. Unión Europea, 2015.

Se están realizando esfuerzos para establecer marcos legales y regulatorios que respalden la Ciencia Abierta, abordando aspectos como la propiedad intelectual, la privacidad de los datos y la ética de la investigación. En este sentido, las políticas de Ciencia Abierta se desarrollan a nivel internacional (por ejemplo, la Recomendación de la UNESCO sobre Ciencia Abierta 2021 y la política de Ciencia Abierta de la Unión Europea); a nivel nacional (La Estrategia Nacional de Ciencia Abierta (ENCA) para el periodo de 2023-2027 en España); y a nivel institucional (como es el caso de la Max Planck Open Access Policy y la Helmholtz Open Science Policy).

A su vez, en España se ha reformado la ley de la ciencia y la innovación que incorpora eficaces medidas para facilitar el acceso abierto a publicaciones científicas y a los datos resultantes de la investigación financiada con fondos públicos.

Asimismo, la nueva ley orgánica del sistema universitario incorpora importantes medidas dirigidas a promover la ciencia abierta al servicio de la sociedad (Abadal, 2023: 8). La implementación de estas leyes y su forma de operar es un asunto que deberá ser analizado, pues influye directamente con los mecanismos para hacer frente a las barreras del acceso abierto dentro de las instituciones.

Pues el fundamento principal de la ciencia abierta radica en la idea de estimar un cambio sistémico en la forma en la manera de hacer ciencia e investigación en los últimos cincuenta años: pasar de las prácticas habituales de resultados de la investigación en publicaciones científicas a compartir y utilizar todos los conocimientos disponibles en una fase más temprana del proceso de investigación (European Commission, 2015: 33).

Por su parte, Palma Peña (2024: 7) afirma que, a la fecha, la ciencia abierta es un tema emergente con una multiplicidad de enfoques, ya que los factores y actores de este movimiento han desarrollado sus principios, funciones y acciones acordes con sus contextos y objetivos.

Por otra parte, Anglada y Abadal (2018: 293) afirman que la ciencia abierta es esto: un cambio de paradigma en la manera de hacer ciencia. Ésta no cambia sustancialmente con respecto a sus motivaciones y objetivos, pero sí lo hace (sustancialmente) en cuanto a sus métodos. El cambio no está en lo que se hace, sino en cómo se hace.

Así pues, un componente central de la ciencia abierta son las normatividades para sustentarla, conformadas por mandatos, declaraciones, políticas públicas, derechos de autor, limitaciones, excepciones y licenciamientos.

3.3. Datos de investigación

Hace más de una década que la temática de los datos de investigación ha tomado notable interés para la comunidad bibliotecaria y la disciplina bibliotecológica y de los estudios de la información. El acceso abierto a la información y la ciencia abierta han sido dos temáticas que han propiciado la aparición de fenómenos relativos a la organización y el acceso a los datos de investigación de una manera abierta y libre de restricciones.

Los datos de investigación pueden concebirse como conjuntos de datos que son utilizados para el desarrollo de investigación y que forman parte de los resultados de dicho proceso indagatorio. Estos datos son de naturaleza y tipología diversa, pueden ser numéricos, textuales o alfanuméricos. Representar a investigaciones de corte cuantitativo y cualitativo.

Childs, Mcleoud y Lomas (2014: 144) afirman que los datos de investigación se abren con dos propósitos: proporcionar evidencia de que la investigación se realizó correctamente y proporcionar datos para su reutilización (análisis secundario) y la generación de nuevos hallazgos y resultados, es decir, “pararse sobre los hombros de gigantes”.

Bajo esta premisa, los datos de investigación no pueden considerarse como abiertos si existe alguna restricción que intervenga en su libre descarga, consulta y disponibilidad. Aunado a ello, se debe considerar que la apertura de datos de investigación se encuentra influenciada por factores intrínsecos y extrínsecos relacionados con el proceso de investigación.

De acuerdo con Inkpen (2021: 4) los puntos clave de los datos abiertos en términos de accesibilidad, escrutinio de la investigación y transparencia, son los valores éticos que todos los científicos deberían aspirar a alcanzar. Se podría argumentar que el escrutinio de la investigación dentro de un marco de racionalismo crítico, un marco que está inherentemente centrado en el análisis crítico y la falsificación/validación, es la manera en que se puede lograr el progreso científico.

En este sentido, la investigación moderna está siendo remodelada por una avalancha de datos, que necesita de metodología y técnicas para obtener procesarlos de una manera sistematizada. En este sentido, “establecer acuerdos de intercambio de datos para colaboraciones complejas e internacionales de investigación en el ambiente digital, requiere políticas nacionales e internacionales, marcos legales y prácticas de gestión de datos apropiados (Fitzgerald, Fitzgerald y Pappalardo, 2009: 207).”

Se estima que las políticas de datos abiertos de investigación reforzarán el principio de reutilización a los datos, un tema que ha despertado la discusión entre las comunidades académicas y científicas a nivel internacional. Pues la reutilización de datos se ha convertido en un nuevo debate académico en el contexto del paradigma de investigación basada en datos, lo cual ha atraído atención en el mundo académico a nivel mundial (Wang, Guang y Lian, 2021: 1161).

Uno de los principales argumentos que se derivan de estas discusiones tiene que ver con los alcances y limitaciones de la reutilización de los datos, pues si bien el mayor potencial de los datos abiertos de investigación al abrirse y reutilizarse recae en su labor como punto de partida para contrastar y validar resultados, por otro lado, existe incertidumbre en la manera en cómo se va a hacer reconocimiento de los autores y creadores de dichos datos.

En términos generales, la ciencia abierta persigue garantizar el acceso abierto tanto a los procesos como a los resultados de la actividad científica mediante la reutilización, la redistribución y la reproducción de las investigaciones (Vitón-Castillo, García-Espinosa y Arencibia-Paredes, 2020: 169).

Pues la reutilización de los datos conlleva efectuar el reconocimiento intelectual a sus generadores. En este sentido, dentro del ecosistema de datos actual, existen principios y estándares que propician este reconocimiento a través del seguimiento de pautas y recomendaciones que tienen el propósito de conformar un entorno holístico de datos, basado en el respeto a la propiedad intelectual.

De acuerdo con Khan, Thelwall y Kousha (2023) se sabe que el intercambio de datos está aumentando en algunas disciplinas para cumplir con los requisitos institucionales y de los organismos de financiación. Sin embargo, los datos de la investigación no siempre se comparten de una manera significativa que pueda conducir a su accesibilidad y reutilización a largo plazo.

De esta manera, compartir y reutilizar datos son conceptos diferentes, pues el segundo se diferencia de los productos que se derivan a partir del análisis y contrastación de los resultados de investigaciones previas en función de los datos que se han utilizado para generar nuevos documentos como artículos, capítulos de libros, conjuntos de datos, libros, disertaciones, etc.

En el caso de los artículos de datos, también conocidos como “data papers”, estos documentos se tratan de un tipo de documento académico que describe un conjunto de datos de manera detallada y sistemática. A diferencia de un artículo de investigación convencional que se centra en los hallazgos y resultados obtenidos a partir de datos, un artículo de datos se enfoca en la descripción exhaustiva de los datos en sí mismos.

Estos documentos suelen incluir detalles sobre la metodología de recopilación, los protocolos de calidad de los datos, la estructura del conjunto de datos, el formato de archivo, las variables incluidas y cualquier otro aspecto relevante para comprender y utilizar efectivamente a los datos. Además, facilitan la reutilización de los conjuntos de datos, promoviendo la colaboración y la maximización del valor de la información recopilada.

3.4. Principios y esquemas para la organización y apertura de los datos de investigación

Actualmente, los datos de investigación se encuentran en un ambiente digital que está caracterizado por el uso de principios y esquemas que tienen el propósito de fomentar el acceso a los datos de una manera sistematizada y organizada. Los repositorios de datos de investigación utilizan interfaces que permiten a los usuarios navegar por las colecciones de datos, en donde se aprecia la interacción de diversos elementos que unifican la capacidad de los sistemas y los usuarios para consultar a dichos datos.

3.4.1. Principios FAIR

Los principios FAIR (Findable=Localizables, Accessible=Accesibles, Interoperable=Interoperables, Reusable=Reutilizables), definen cuales son las características que los recursos, herramientas, vocabularios e infraestructuras de datos contemporáneos deben contemplar para ayudar al descubrimiento y reutilización de los datos por parte de terceros, es decir consumidores y usuarios de datos (Wilkinson, et al. 2016: 4).

Para cumplir con estas características, los principios FAIR sustentan su funcionamiento bajo el contexto de los metadatos.

Box 2 | The FAIR Guiding Principles

To be Findable:
 F1. (meta)data are assigned a globally unique and persistent identifier
 F2. data are described with rich metadata (defined by R1 below)
 F3. metadata clearly and explicitly include the identifier of the data it describes
 F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:
 A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 A1.1 the protocol is open, free, and universally implementable
 A1.2 the protocol allows for an authentication and authorization procedure, where necessary
 A2. metadata are accessible, even when the data are no longer available

To be Interoperable:
 I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
 I2. (meta)data use vocabularies that follow FAIR principles
 I3. (meta)data include qualified references to other (meta)data

To be Reusable:
 R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 R1.1. (meta)data are released with a clear and accessible data usage license
 R1.2. (meta)data are associated with detailed provenance
 R1.3. (meta)data meet domain-relevant community standards

(fig.2) Principios FAIR. Fuente: The FAIR Guiding Principles for scientific data management and stewardship | Scientific Data, sin fecha [en línea]. Recuperado a partir de: <https://www.nature.com/articles/sdata201618> [accedido 8 noviembre 2023].

En la figura 2, puede apreciarse como interactúan los metadatos con cada uno de los principios FAIR. En este sentido, los metadatos desempeñan un papel crucial en la capacidad de encontrar datos. Proporcionan información sobre los datos, como su origen, contexto, formato y contenido. Al utilizar metadatos estructurados y descriptivos, se facilita la búsqueda y descubrimiento de datos.

De acuerdo con Sansone, et, al. (2019), como ocurre con cualquier otro objeto digital, los estándares, las bases de datos y los repositorios son de naturaleza dinámica, con un 'ciclo de vida' que abarca la formulación, el desarrollo y el mantenimiento; su estatus en este ciclo puede variar dependiendo del nivel de actividad del grupo o comunidad en desarrollo. Por lo que los principios FAIR son inherentes a este ciclo y su implementación continua.

3.4.2. Open Aire

Se trata de un proyecto tecnológico creado por la Comisión Europea con el objetivo de apoyar, acelerar y medir la correcta implementación de las políticas europeas de acceso abierto a las publicaciones científicas, datos de investigación y software (OpenAire, 2023).

OpenAire ha pasado por cinco fases de desarrollo, lo que le ha permitido apoyar diversos programas y proyectos para fomentar el progreso de la ciencia abierta en Europa. La fase actual de OpenAIRE, denominada Nexus, se centra en mantener y mejorar la estructura de servicios para el apoyo de la Ciencia Abierta en todas sus fases: publicar investigaciones, reutilizar datos, supervisar su repercusión y ayudar a fomentar sus hallazgos.

Los componentes de la fase Nexus son esenciales para fomentar la interoperabilidad entre los datos de investigación y la agregación de diversas plataformas de información científica y académica. Se trata de unificar los desarrollos basados en ciencia abierta y comunicarlos interoperablemente mediante el uso de estándares de metadatos.

3.4.3. Datacite Metadata Schema

El uso de metadatos sirve para propiciar la organización y accesibilidad de los datos. Proporcionan detalles sobre la ubicación de los datos, los requisitos para acceder a ellos y las condiciones de uso. Los metadatos bien definidos ayudan a garantizar que los datos sean accesibles de manera clara y comprensible.

Bajo esta premisa, Datacite Metadata Schema es una lista de propiedades principales de metadatos elegidas para obtener información precisa y consistente, para la identificación de un recurso de información con fines de citación y recuperación, mediante instrucciones de uso recomendadas en documentación (Datacite Metadata Working Group, 2021).

Este esquema de metadatos proporciona elementos para la descripción y representación de conjuntos de datos de investigación, además de presentar un marco común para la citación de dichos datos mediante el uso de referencias y documentación que explica el contexto de los datos y su naturaleza.

3.4.4. Data Documentation Initiative (DDI)

Es un estándar internacional de metadatos para la descripción de datos de investigación que son producidos en los ámbitos de las ciencias sociales, las humanidades, las ciencias del comportamiento, las ciencias económicas y ciencias de la salud (Data Documentation Initiative, 2023). DDI es un estándar que permite documentar y gestionar diferentes etapas del ciclo de vida de los datos de investigación, como su organización, recopilación, procesamiento, distribución, descubrimiento y archivo mediante un proceso global de sistematización.

Este esquema se encuentra implementado en diferentes iniciativas y plataformas para la organización y sistematización de los datos de investigación, por ejemplo, el proyecto Dataverse de Harvard, el catálogo de datos del Consorcio de Archivos Europeos de Datos de Ciencias Sociales (CESSDA), el Archivo Australiano de Ciencias Sociales, entre otros. Con este esquema es posible describir y representar los atributos de los conjuntos de datos de investigación mediante la creación de registros descriptivos que responde a un perfil de metadatos desarrollado exprefeso para esta tarea.

Dataverse Add Data - Search - About User Guide Support Eder Ávila

Datos correspondientes a registros bibliográficos de la colección general de la biblioteca del IIBI

Draft Unpublished

Avila, Eder, 2023, "Datos correspondientes a registros bibliográficos de la colección general de la biblioteca del IIBI", <https://doi.org/10.70122/FK2/9J5HB0>, Demo Dataverse, DRAFT VERSION, UNF:6:KGnxugik5Uru8/yypMGkzQ== [fileUNF]

Cite Dataset - Learn about [Data Citation Standards](#).

Access Dataset -
 Publish Dataset -
 Edit Dataset -
 Contact Owner Share

Description Conjunto de datos en formatos csv y rdf relativos a registros bibliograficos de la Biblioteca del IIBI. Se utilizaron para el desarrollo de un Grafo RDF para la generación de una consulta con SPARQL. (2023-12-13)

Subject Arts and Humanities

Keyword Bibliographic Data, Linked Open Data, Knowledge graph

Related Publication Ávila-Barrientos, Eder. 2022. "Recuperación de información con Linked Open Data". Investigación Bibliotecológica: archivonomía, bibliotecología e información 36 (91): 125-146. doi: DOI

License/Data Use Agreement CC0 1.0

Files Metadata Terms Versions

Add + Edit Metadata

Citation Metadata

Persistent Identifier	doi:10.70122/FK2/9J5HB0
Title	Datos correspondientes a registros bibliográficos de la colección general de la biblioteca del IIBI
Author	Avila, Eder (UNAM) - ORCID: https://orcid.org/0000-0001-5439-0495
Point of Contact	Use email button above to contact. Avila, Eder (UNAM)
Description	Conjunto de datos en formatos csv y rdf relativos a registros bibliograficos de la Biblioteca del IIBI. Se utilizaron para el desarrollo de un Grafo RDF para la generación de una consulta con SPARQL. (2023-12-13)
Subject	Arts and Humanities
Keyword	Bibliographic Data Linked Open Data Knowledge graph
Related Publication	Ávila-Barrientos, Eder. 2022. "Recuperación de información con Linked Open Data". Investigación Bibliotecológica: archivonomía, bibliotecología e información 36 (91): 125-146. doi: DOI http://dx.doi.org/10.22201/iibi.24488321xe.2022.91.58567
Language	Spanish, Castilian
Depositor	Avila, Eder
Deposit Date	2023-12-13
Software	Open Refine, Version: 3.4 GraphDB, Version: 9.1

(fig.3) Registro de metadatos de un conjunto de datos de investigación. Elaboración propia mediante el uso de Dataverse, 2023.

Como se puede apreciar en la figura 3, los metadatos son esenciales para representar a los atributos descriptivos y de contenido que se encuentran en los conjuntos de datos de investigación, además de propiciar el debido reconocimiento de la propiedad intelectual de los datos, lo cual aporta elementos para la latente reutilización de los datos. En suma, los metadatos son un medio por el cual los datos de investigación pueden visibilizarse y recuperarse para su consulta en un entorno sistematizado como puede ser un repositorio de datos.

4. Resultados

Tabla 2 Matriz de análisis comparativo de repositorios de datos de investigación

VARIABLE	R3DATA	DATAVERSE	ZENODO	MENDELEY DATA	DRYAD	CORA	E-Ciencia Datos
Breve descripción	Registro global de repositorios de datos de investigación multidisciplinarios es financiado por la Fundación Alemana de Investigación	Repositorio de datos gratuito abierto a todos los investigadores de cualquier disciplina, tanto dentro y fuera de la comunidad de Harvard, donde puedes compartir, archivar, citar, acceder y explorar datos de investigación	Proyecto financiado por el CERN, ofrece una plataforma de acceso abierto y gratuito para la publicación y preservación de conjuntos de datos multidisciplinarios	Repositorio especializado para datos de investigación. Recopila y comparte conjuntos de datos con la comunidad de investigación siguiendo principios FAIR.	Plataforma de publicación de datos abiertos y una comunidad comprometida con la disponibilidad abierta y la reutilización rutinaria de todos los datos de investigación. Todos los datos de Dryad se someten a un proceso de curación y se publican bajo una exención CC0 para promover su reutilización.	Repositorio de datos federado y multidisciplinar que permite a las universidades catalanas, centros de investigación CERCA y otras entidades de investigación publicar conjuntos de datos de investigación de forma JUSTA y siguiendo las directrices de la EOSC.	cienciaDatos, como infraestructura proporcionada por el Consorcio Madroño, es garantizar, en el contexto de la ciencia abierta, el depósito, la preservación y la difusión de los datos de investigación generados por las instituciones miembro de acuerdo con los principios FAIR
Temática	Multidisciplinario	Multidisciplinario	Multidisciplinario	Multidisciplinario	Multidisciplinario	Multidisciplinario	Multidisciplinario
Límites de tamaño de conjuntos de datos	Ninguno	Sin límite de tamaño de bytes por conjunto de datos. Establece un límite de tamaño de archivo de 2,5 GB.	50 GB por conjunto de datos	10 GB por conjunto de datos	300 GB por conjunto de datos	100 GB por conjunto de datos	50 GB por conjunto de datos
Espacio de almacenamiento por investigador	Sin limite	1 TB por investigador	Sin limite	Sin limite	Sin limite	Sin limite	Sin limite
Identificador único y persistente que utiliza	DOI HANDLE	DOI HANDLE	DOI	DOI	DOI	DOI	DOI
Opciones de licencia	CC COPYRIGHTS CC0 APACHE LICENSE2.0 PUBLIC DOMAIN ODC BSD	Por defecto, los conjuntos de datos se publican en el dominio público (CC0). Los depositantes pueden cambiar el uso de licencias	El contenido está disponible públicamente bajo cualquiera de las 400 licencias abiertas (desde opendefinition.org y spdx.org). También se admite contenido restringido y cerrado.	Licencias predeterminadas: CC0 1.0 CC BY 4.0 CC BY NC 3.0 MIT Apache 2.0 BSD 3-Clause BSD 2-Clause GPL v3 GPL v2 LGPL MPL-2.0 CeCILL CeCILL-B CERN OHL TAPR OHL	CC0	CC CC0	CC CC0

Costos para el investigador	Gratuito	Gratuito (hasta 1 TB por investigador)	Gratuito hasta 50 GB por conjunto de datos	Gratuito	Gratuito	Gratuito	Gratuito para investigadores del Consorcio Madroño
Esquemas de metadatos admitidos	Dublin Core DataCite DDI Darwin Core ISO-19115	Dublin Core, DDI DataCite Schema.org OAI	DataCite	Dublin Core, DataCite, Schema.org	DataCite, schema.org	DDI	DDI
Formatos admitidos para exportar	JSON XML RDF	JSON XML	DataCite, Dublin Core, DCAT-AP, JSON, JSON-LD, GeoJSON, MARCXML	JSON, XML	JSON, XML, schema.org	CSV XML	CSV XML
Soporte para identificadores de creadores/autores	ORCID ISNI AUTHORCLAIM RESEARCHERID	ORCID, ISNI, LCNA, VIAF, GND, DAI, ResearcherID and ScopusID	ORCID	ORCID - Mendeley ID - Scopus Author ID	ORCID	ORCID	ORCID
Admite enlaces a publicaciones relacionadas	Si	Si	DataCite tipos de relación	Sí, DataCite y Scholix tipos de relación	DataCite tipos de relación	Si	Si
Vinculación de productos derivados	Si	Si	Si	Sí, DataCite y Scholix tipos de relación	DataCite tipos de relación	Si	Si
Afiliaciones de creador/autor	Si	Si	Si	Si	Si	Si	Si
Enlace a software relacionado	Si	Si	Si	Si	Si	Si	Si
Enlace a un producto de investigación relacionado	Si	Si	DataCite tipos de relación	Si, mediante relaciones semánticas	DataCite tipos de relación	Si	Si
Métricas de uso de datos admitidas	Descargas, visualizaciones, citas	Descargas, visualizaciones, citas	Descargas, visualizaciones, citas, altimetrías	Descargas, visualizaciones, citas, altimetrías	Descargas, visualizaciones, citas, altimetrías	Descargas, visualizaciones, citas	Descargas, visualizaciones, citas
Soporte de Embargo	Si	Si	Si	Si	Si	Si	Si
Admite la revisión por pares durante el embargo (por ejemplo, "enlace para compartir temporal")	Si	Si	Si	Si	Si	Si	Si
Admite acceso administrado	Si	Si	Si	Si	Si	Si	Si
Preservación	Si a través de estándares de preservación	Si, a través de estándares de preservación	Estrategia CERN Archive	Estrategia DANS Archive	Estrategia DANS Archive	Estrategia de preservación de mínimo 10 años	Plan de preservación de e-Ciencia Datos
Enlaces a registros para más información	Si	Si	Si	Si	Si	Si	Si
Registro FAIRsharing	https://doi.org/10.25504/FAIRsharing.fae0b7	https://doi.org/10.25504/FAIRsharing.t2e1s	https://doi.org/10.25504/FAIRsharing.wy4eqf	https://doi.org/10.25504/FAIRsharing.3epmpp	https://doi.org/10.25504/FAIRsharing.wkggtx	https://doi.org/10.25504/FAIRsharing.av9fslp	https://doi.org/10.25504/FAIRsharing.lone3q

Registro Re3data	https://www.re3data.org/	https://www.re3data.org/ repository/r3d100010051	https://www.re3data.org/ repository/r3d100010468	https://www.re3data.org/ repository/r3d100011868	https://www.re3data.org/ repository/r3d100000044	https://www.re3data.org/ repository/r3d100013559	https://www.re3data.org/ repository/r3d100012316
---------------------	---	--	--	--	--	--	--

Tabla de elaboración propia con base a Stall, et al. 2020.

El resultado principal de este trabajo ha consistido en el desarrollo de un análisis comparativo de los repositorios de datos de investigación, lo que ha permitido obtener una visión holística de las variables de los repositorios en función de los atributos que han sido utilizados para describirlos y analizarlos mediante la matriz que se presenta en la tabla 2. Si bien, una comparación entre repositorios puede llevarla a cabo el personal bibliotecario o gestor de una determinada institución académica, es importante resaltar que las variables de comparación resumidas en la matriz fueron recabadas a partir del estudio de las propiedades de los repositorios y de los datos de investigación por sí mismos, por lo que la matriz puede representar una fuente de referencia para llevar a cabo una comparación metodológica de repositorios de datos de investigación.

Como puede observarse las variables permiten conocer las propiedades descriptivas y representativas de los repositorios de una manera integral, pues cada uno de los atributos se relaciona con el análisis estructural de los repositorios en concordancia con los elementos señalados en el discurso de la Ciencia Abierta.

De esta manera se tomaron en cuenta variables como la descripción del repositorio, su temática, los límites de tamaño de los conjuntos de datos que pueden almacenar, el espacio de almacenamiento por investigador, el uso de identificadores únicos y persistentes, las opciones de licencia que aplican, los costos para el investigador, los esquemas de metadatos admitidos, los formatos admitidos para exportar datos y el soporte para identificadores de creadores/autores.

Además se consideraron propiedades como la admisión de enlaces a publicaciones relacionadas con los datos, la vinculación de productos derivados, afiliaciones de creador/autor de los datos, el uso de enlaces a software relacionado, enlace a un producto de investigación relacionado, empleo de métricas de uso de datos, soporte para periodos de embargo, admisión de revisión por pares durante el embargo (por ejemplo, "enlace para compartir temporal"), admisión de acceso administrado, preservación de enlaces a registros para más información, registro FAIRsharing y registro Re3data.

Tomando como base la metodología propuesta por Stall, et al. (2020) se analizaron repositorios de datos de investigación con diferentes objetivos y estructuras, esto con la intención de comparar e identificar las similitudes y diferencias entre los atributos de cada repositorio acorde a los elementos que se enmarcan en el contexto de la Ciencia Abierta.

Los resultados obtenidos, permiten apreciar la interoperabilidad entre los repositorios analizados, pues la mayoría de ellos hacen uso de los principios FAIR para respaldar el funcionamiento de los repositorios con base a los términos de acceso y reutilización de los datos de investigación.

En este sentido, los resultados permiten obtener un acercamiento a los elementos que deben considerarse para la implementación de repositorios de datos de investigación. Es importante destacar el papel que el repositorio CORA y E-Ciencia Datos tienen en el fomento de la ciencia abierta en el contexto español pues son dos repositorios que aplican política y fundamentos para la reutilización de datos y la visibilidad de los resultados de investigación.

Como parte de los resultados, se destaca el carácter multidisciplinario de los repositorios estudiados, pues es posible consultar datos de diversas disciplinas como la medicina, la demografía, las ciencias sociales y las humanidades. Esto refleja que los datos que se concentran en estos repositorios son de carácter y temática ampliamente diversa.

La mayoría de los repositorios establece límites de tamaño de los conjuntos de datos que pueden almacenar, se destaca la amplia capacidad de 300 Gb que Dryad maneja en contraposición a los 10 Gb por conjunto de datos que establece Mendeley Data. En esta categoría llama la atención el papel de Re3data pues al tratarse de un directorio de repositorios, su función recae en tratarse de un instrumento referencial y listar a repositorios, más que concentrarse en almacenar conjuntos de datos perse.

A su vez, solo Dataverse establece límites en el espacio de almacenamiento por parte de los investigadores, siendo una constante que todos los demás repositorios no establecen límites para este cometido. Por su parte, puede observarse el uso de diferentes identificadores únicos y persistentes siendo más reiterativo el uso de los Identificadores de Objetos Digitales (DOI). En cuanto al uso de licencias, se puede apreciar una amplia variedad de instrumentos, cabe destacar el uso de la licencia CC0 para el caso de los datos que son depositados en los repositorios, y el uso de licencias específicas para cada tipo de publicación y producto que se hace referencia mediante los datos que han sido sistematizados, por ejemplo, el caso de Creative Commons en diferentes niveles.

La mayoría de los repositorios son de carácter gratuito con excepción de Dataverse y Zenodo, los cuales manejan cuotas dependiendo el tamaño de los conjuntos de datos que se deseen almacenar. En cuanto al uso de esquemas de metadatos, DataCite y DDI son los de uso específico para la organización y descripción de conjuntos de datos de investigación, no obstante, se observa la adaptabilidad de Dublin Core para representar las propiedades de los datos de investigación, a pesar de que es un esquema que no ha sido desarrollado para este propósito.

Se observa una constante en cuanto al uso de JSON (formato de texto para el intercambio de datos) como formato para la exportación de datos dentro de los repositorios, pues se trata de un formato que es flexible y compatible con la mayoría de los servicios de intercambio de datos, además el uso de CSV (valores separados por comas) y XML (lenguaje de marcado extensible) forman parte también de esta tendencia.

El uso de ORCID como identificador persistente para autores y creadores de datos tiene presencia en todos los repositorios analizados, además se incluyen otros identificadores como ISNI lo que permite establecer relaciones entre los datos, sus publicaciones y los datos que generan.

En lo que se refiere a la admisión de enlaces a publicaciones relacionadas, la vinculación de productos derivados, las afiliaciones de creador/autor y el enlace a software relacionado con la generación o consulta de los datos, son variables que forman parte de una constante en los repositorios analizados.

En todos los repositorios se emplean métricas de uso de datos como las descargas realizadas, las visualizaciones y citaciones. Esta variable permite obtener una valoración de los datos en función de su acceso, consulta y uso por un determinado usuario.

A su vez, los repositorios analizados manejan ayudas y soporte para explicar el periodo de embargo que puede estar aplicado en diversos conjuntos de datos, además de ofrecer acceso administrado a dichos conjuntos, lo cual tiene repercusión en la disponibilidad y accesibilidad de los datos en un determinado periodo de tiempo.

En cuanto al atributo de preservación cabe destacar la generación de planes y estrategias de preservación en repositorios como CORA y E-Ciencia Datos pues se trata de instrumentos que fundamenta el ciclo de vida de los datos de investigación en cuanto a su uso y relevancia se refiere. Además, los repositorios utilizan estándares de preservación y estrategias europeas como parte de este propósito tal es el caso de DANS Archive.

Todos los repositorios analizados emplean enlaces a registros para ampliar información acerca de un detalle específico de los datos, además de estar registrados en plataformas como Re3data y FAIRSharing. Esta particularidad permite observar la generación de un ecosistema de datos de investigación, caracterizado por el uso de enlaces y registros que intentan propiciar una interoperabilidad para el intercambio y vinculación de datos de investigación.

Por otra parte, los resultados de este análisis revelan que los repositorios de datos de investigación difieren en una serie de aspectos, incluyendo:

- Área temática: Los repositorios de datos de investigación pueden centrarse en una sola área temática o en múltiples áreas temáticas.
- Disciplina científica: Los repositorios de datos de investigación pueden centrarse en una sola disciplina científica o en múltiples disciplinas científicas.
- Tamaño: Los repositorios de datos de investigación pueden variar en tamaño, desde repositorios pequeños que albergan unos pocos miles de conjuntos de datos hasta repositorios grandes que albergan millones de conjuntos de datos.
- Formato de los datos: Los repositorios de datos de investigación pueden admitir una variedad de formatos de datos, incluyendo datos estructurados, datos no estructurados y datos semiestructurados.
- Licencias de los datos: Los datos de los repositorios de datos de investigación pueden estar disponibles bajo una variedad de licencias, incluyendo licencias de acceso abierto, licencias de uso comercial y licencias propietarias.
- Características de los conjuntos de datos: Los conjuntos de datos de los repositorios de datos de investigación pueden variar en función de una serie de características, incluyendo la calidad de los datos, la cantidad de datos, la antigüedad de los datos y la relevancia de los datos para los investigadores.

En función de estos criterios, los resultados del análisis comparativo de repositorios de datos de investigación pueden revelar que algunos repositorios son más adecuados para determinados tipos de datos o para determinados tipos de investigadores. Por ejemplo, un repositorio que se centra en una sola área temática puede ser más adecuado para investigadores que trabajan en esa área temática específica.

Un repositorio que admite un formato de datos específico puede ser más adecuado para investigadores que trabajan con ese formato de datos. Y un repositorio que ofrece licencias de uso comercial puede ser más adecuado para investigadores que necesitan utilizar los datos para fines exhaustivos.

Un análisis comparativo de repositorios de datos de investigación en ciencias de la salud puede revelar que algunos repositorios se centran en datos clínicos, mientras que otros se centran en datos de investigación básica.

Asu vez, un análisis comparativo de repositorios de datos de investigación en ciencias sociales puede revelar que algunos repositorios se centran en datos de encuestas, mientras que otros se centran en datos de observación. Por su parte, un análisis comparativo de repositorios de datos de investigación en ciencias de la Tierra puede arrojar que algunos repositorios se centran en datos de un determinado satélite, mientras que otros se centran en datos de campo.

El dinamismo que se percibe en los repositorios de datos de investigación está fundamentado en el origen y disponibilidad de los datos en diversos contextos digitales. En este trabajo se destaca la interoperabilidad que se pretende establecer entre los datos de investigación mediante el uso de estándares, políticas y licenciamientos relativos al movimiento de la Ciencia Abierta.

5. Discusión

A través de este trabajo, se plantea la importancia de desarrollar estándares para la descripción e intercambio de datos científicos de forma interoperable. Si bien la organización y sistematización de datos es una realidad, su comunicación efectiva sigue siendo un desafío, a pesar de la existencia de pautas como los registros R3Data y FAIRSharing para su vinculación y referenciación.

- Formatos de datos: Definición de formatos de datos estandarizados para facilitar la interoperabilidad y el intercambio de datos entre diferentes repositorios.
- Metadatos: Establecimiento de estándares para la descripción de metadatos que incluyan información relevante sobre los datos, como su origen, metodología de recolección, variables y condiciones de uso.
- Identificadores únicos: Asignación de identificadores únicos persistentes a los conjuntos de datos para facilitar su referencia y citación.
- Licencias: Definición clara de las licencias de uso de los datos para permitir su reutilización responsable.

La adopción de estándares interoperables por parte de la comunidad científica impulsaría una mayor colaboración, transparencia y acceso a los datos de investigación, acelerando el ritmo del progreso científico.

Si bien la matriz de análisis de repositorios de datos de investigación ofrece un conjunto de variables para comparar estos repositorios, es importante reconocer que no se trata de una herramienta exhaustiva. La naturaleza dinámica y diversa de los datos de investigación, con sus características en constante evolución, presenta un reto significativo para capturar todas sus singularidades en un estudio de este tipo.

Se estima que, para los investigadores, un análisis de este tipo puede ayudar a identificar el repositorio más adecuado para sus necesidades. Los investigadores pueden utilizar los resultados de un análisis comparativo para comparar repositorios en función de criterios como el área temática, la disciplina científica, el tamaño, el formato de los datos, las licencias de los datos y las características de los conjuntos de datos.

Por otra parte, los bibliotecarios, de bibliotecas académicas y especializadas, pueden tomar como punto de referencia el análisis comparativo efectuado en este trabajo. En donde contemplen las variables para identificar cual repositorio de datos de investigación puede ser susceptible de utilizarse para dar mayor visibilidad a los datos generados por las comunidades académicas y de investigación. Según Rauf y Du (2018), “las bibliotecas académicas están atravesando un período de intenso cambio, en el que se están creando nuevos roles y servicios para enfrentar el desafío de los macrodatos. Frente a las crecientes demandas de las agencias de financiamiento, el personal docente, los estudiantes y los investigadores, los bibliotecarios académicos necesitan crear y apoyar investigaciones intensivas en datos”.

De esta manera, el estudio desarrollado en este trabajo también está dirigido a aquellos bibliotecarios expertos en el procesamiento de datos, pues en el panorama actual de la investigación, donde la apertura, la colaboración y el intercambio de conocimiento son pilares fundamentales, el procesamiento de datos de investigación ha adquirido una importancia crucial. En este contexto, los bibliotecarios se erigen como actores clave, desempeñando un papel fundamental en este procesamiento y estableciendo una relación simbiótica con la comunidad científica.

Por otra parte, de acuerdo con lo señalado por Abadal (2023) en el más reciente informe de Ciencia Abierta en España, los investigadores resaltan la importancia que adquieren los datos por la transparencia y visibilidad que ofrecen a los resultados de investigación, y también el posibilitar la validación de estos, sin menospreciar su potencial reutilización. Sin embargo, existe todavía un alto desconocimiento de cómo gestionar y compartir los datos, y preocupación por los posibles malos usos de los datos.

A pesar de estos desafíos, los repositorios de datos de investigación son una herramienta importante para el avance de la Ciencia Abierta y su evolución, pues estos repositorios desempeñarán un papel cada vez más importante en el acceso y la transparencia de los resultados de la investigación científica.

En el contexto de la Ciencia Abierta, un análisis comparativo de repositorios de datos de investigación puede ayudar a promover la reutilización de los datos de investigación. La reutilización de los datos de investigación puede conducir a nuevos descubrimientos científicos, a la reducción del desperdicio de recursos y a la democratización del conocimiento científico y a la conformación de un ecosistema interoperable de datos de investigación.

Al respecto Sansone (2019: 359) afirma que citar un registro de FAIRsharing para un estándar, base de datos y repositorio ofrece una vista rápida de todos los descriptores e indicadores relacionados con un recurso, así como cualquier evidencia de adopción o respaldo por parte de una política u organización de investigación.

Sin embargo, es importante considerar las limitaciones de este tipo de análisis. Primeramente, los criterios de comparación utilizados varían según las necesidades de los investigadores, las instituciones académicas o las organizaciones de financiación de la investigación.

En segundo lugar, los datos de los repositorios de datos de investigación pueden cambiar con el tiempo, por lo que es importante que los investigadores verifiquen los resultados de un análisis comparativo antes de utilizarlos para tomar una decisión.

A pesar de estas limitaciones, se estima que los análisis comparativos de repositorios de datos de investigación pueden ser una herramienta valiosa para promover la ciencia abierta y la reutilización de los datos.

6. Conclusiones finales

La era digital y el tsunami de datos presentes en la actualidad son factores que han impulsado el análisis comparativo de repositorios de datos de investigación, revelando características y tendencias esenciales para el avance de la Ciencia Abierta, pues estos repositorios son elementos clave para compartir y reutilizar datos de manera responsable y ética, promoviendo la transparencia y el progreso científico.

Los repositorios de datos de investigación experimentan un crecimiento exponencial en el ámbito académico y científico, impulsados por su implementación en universidades e instituciones de investigación. No obstante, para que su adopción sea efectiva y sostenible, resulta indispensable que se asiente sobre sólidos cimientos de Ciencia Abierta. Esto implica la internalización de sus principios por parte de todos los actores involucrados en el ecosistema de datos actual.

La disparidad observada en la implementación de estos repositorios a nivel global pone de manifiesto una brecha evidente en el uso y explotación de datos entre diferentes países del mundo. Abordar esta brecha requiere un esfuerzo conjunto por parte de la comunidad científica, las instituciones académicas, los gobiernos y las organizaciones internacionales para fomentar el desarrollo de políticas de Ciencia Abierta que promuevan el intercambio abierto y responsable de datos.

En este sentido, los investigadores reconocen la importancia de compartir datos para la validación, visibilidad y reutilización de sus resultados de investigación. Sin embargo, existen dudas sobre las mejores prácticas para la gestión y el intercambio de datos, especialmente en cuanto a su uso indebido se refiere.

En este sentido, resulta fundamental la participación de los bibliotecarios en el proceso de gestión de datos de investigación, ya que pueden convertirse en aliados indispensables para los investigadores en este proceso. Su experiencia, conocimiento y habilidades les permiten contribuir a la organización, preservación, acceso y reutilización efectiva de los datos, impulsando el avance científico y una mayor visibilidad del conocimiento en la era de la Ciencia Abierta.

7. Agradecimientos

Investigación realizada gracias al Programa UNAM-DGAPA-PASPA 2023

8. Referencias bibliográficas

- Abadal, E. et al [2023]. *Ciencia abierta en España 2023: informe de situación y análisis de la percepción*. Barcelona-València: Grupo Ciencia Abierta. Disponible en: <http://hdl.handle.net/2445/200020>
- Anglada, L.; Abadal, E. [2018]. ¿Qué es la ciencia abierta? *Anuario ThinkEPI*, 12, 292-298. <https://doi.org/10.3145/thinkepi.2018.4>
- Bautista J., Badillo, R. y Lobato, C. [2023]. *Ciencia abierta como una nueva forma de hacer investigación*, 26, 1-23. Disponible en: <https://www.redalyc.org/journal/5534/553476030008/html/>
- Childs, S. et al. [2014] Opening research data: issues and opportunities. Anne Thurston. Editora, *Records Management Journal*. 24 (2), 142-162. Disponible en: 10.1108/RMJ-01-2014-0005
- Data Documentation Initiative. About*, [2023]. Disponible en: <https://ddalliance.org/>
- Datacite metadata working group. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data and Other Research Outputs*, 2021. Version 4.4. DataCite e.V. Disponible en: <https://doi.org/10.14454/3w3z-sa82>.
- European Commission, Open innovation, open science, open to the world: a vision for Europe* [2015]. Luxemburgo: Publications Office of the European Union. ISBN 978-92-79-57346-0. Disponible en: <https://data.europa.eu/doi/10.2777/061652> [Consulta 24 de noviembre 2023].
- Fitzgerald, A., Fitzgerald, B. y Pappalardo, K. [2009] The future of data policy. En: *The fourth paradigm: Data intensive scientific discovery*. Tony Hey, Stewart Tansley y Kristin Tolle (eds). USA: Microsoft Research, 201-208. Disponible en: http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_part4_fitzgerald.pdf
- Inkpen, R., Gauci, R. y Gibson, A. [2021]. The values of open data. *Area*. 53 (2), 240-246. Disponible en: <http://10.1111/area.12682>
- Khan, A., et, al.[2024]. Global overview of research data repositories: an analysis of re3data registry, *Information Discovery and Delivery*, 52 (1), 53-61. Disponible en: <https://doi.org/10.1108/IDD-07-2022-0069>
- Miedema, F. [2022]. *Open Science: The very idea*. Netherlands: SpringerLink, Disponible en: <https://library.oapen.org/bitstream/handle/20.500.12657/51498/9789402421156.pdf?sequence=1&isAllowed=1>
- Open Aire. About, 2023. Disponible en: <https://www.openaire.eu/about>
- Open Science and Research Initiative. *The Open Science and Research Handbook* v1. Disponible en: <https://www.fosteropenscience.eu/content/open-science-and-research-handbook>
- Paic, A. Open Science [2021]: Enabling Discovery in the Digital Age, *Going Digital Toolkit Note*, 13, Disponible en: https://goingdigital.oecd.org/data/notes/No13_ToolkitNote_OpenScience.pdf.
- Palma Peña, J., [2024]. Ciencia abierta y normatividades para implementar su apertura en Latinoamérica: tendencias, análisis y recomendaciones. *e-Ciencias de la Información*. DOI 10.15517/eci.v14i1.55664.
- Pampel, H. y Dallmeier-Tiessen, S. [2014]. Open Research Data: From Vision to Practice. In: Bartling, S., Friesike, S. (eds) *Opening Science*. Cham: Springer, Disponible en: https://link-springer-com.pbidi.unam.mx/2443/chapter/10.1007/978-3-319-00026-8_14
- Rauf Khan, H. y Yunfei Du. [2018]. *What is a Data Librarian?: A Content Analysis of Job Advertisements for Data Librarians in the United States Academic Libraries*. Session 139 - Data Librarian: Needs and Qualifications - Education and Training, Science and Technology, and Continuing Professional Development and Workplace Learning. IFLA WLIC 2018 – Kuala Lumpur, Malaysia – Transform Libraries, Transform Societies. Disponible en: <https://library.ifla.org/id/eprint/2255/>
- Sansone, SA., Mcquilton, P., Rocca-Serra, P. et al.[2019]. FAIRsharing as a community approach to standards, repositories and policies. *Nat Biotechnol*, 37, 358–367. Disponible en: <https://doi.org/10.1038/s41587-019-0080-8>
- Stall, S., Martone, M. E., Chandramouliswaran, I., et al. [2020]. *Generalist Repository Comparison Chart*. Zenodo. Disponible en: <https://doi.org/10.5281/zenodo.3946720>
- Suber, P. [2015]. *Acceso Abierto*. México, Universidad Autónoma del Estado de México, ISBN: 978-607-422-627-0
- The State of Open Data 2023. *Digital Science Report*. Disponible en: <https://doi.org/10.6084/m9.figshare.24428194.v1>
- UNESCO. *Recomendación de la UNESCO sobre la Ciencia Abierta*, 2023 [en línea]. Disponible en: https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa. [Consulta: 17 de enero de 2024].

- Vitón-Castillo AA, García-Espinosa E, Arencibia-Paredes NM [2020]. Bases para la implementación de la ciencia abierta. *Revista Información Científica*, 99 (2), 168-177. Disponible en: <http://www.revinfcientifica.sld.cu/index.php/ric/article/view/2890>
- Wang, X., Duan, Q. y Liang, M.[2021], Understanding the process of data reuse: An extensive review. *Journal of the Association for Information Science and Technology*. 72 (9), 1161-1182. Disponible en: <https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.24483>
- Wilkinson, M. et al. [2016]. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3. Disponible en: <https://doi.org/10.1038/sdata.2016.18>