



Recibido: 2 de mayo de 2025
Revisado: 20 de junio de 2025
Aceptado: 29 de junio de 2025

Dirección de los autores:

Facultad de Ingeniería. Universidad
de la República Uruguay (UdelaR),
Avenida Julio Herrera y Reissig 565,
11300 Montevideo (Uruguay)

E-mail / ORCID

gonzalo.tortero@fing.edu.uy

 <https://orcid.org/0009-0003-5004-0342>

rmotz@fing.edu.uy

 <https://orcid.org/0000-0002-1426-562X>

ARTÍCULO / ARTICLE

Datos abiertos en investigación educativa: ¿cumplen los repositorios su función?

Open Educational Research Data: Are Repositories Delivering?

Gonzalo Torterolo-Silva y Regina Motz

Resumen: En el contexto digital actual, que tiene al movimiento de la ciencia abierta como un nuevo paradigma de producción y circulación de conocimiento, este trabajo analiza en qué medida los repositorios de datos de investigación responden a las necesidades técnicas, legales y éticas de los investigadores en educación y ciencias sociales. El abordaje se realiza aplicando un enfoque que combina tres actividades: revisión bibliográfica, entrevista a experto y análisis de catálogos de repositorios. De la fusión de los resultados obtenidos, se concluye que las características de los repositorios que resultan más valoradas por los investigadores de estas disciplinas son la reputación de los repositorios, facilidad de uso, disponibilidad de información de contexto, sostenibilidad (conservación de los datos a largo plazo), apoyo técnico y cumplimiento de los principios FAIR (que promueven la producción de datos Encontrables, Accesibles, Interoperables, Reutilizables). Además, este artículo profundiza en el rol que cumplen los catálogos de repositorios para facilitar la selección de repositorios adecuados para depositar los datos, proporcionando a los investigadores información útil para evaluar la calidad de los repositorios, en cuanto a si responden a sus necesidades. Este estudio proporciona información valiosa para promover la adopción de prácticas de datos abiertos en el campo de la investigación educativa.

Palabras clave: Herramientas de investigación, Archivos, Catálogos online, Metadatos, Uso de datos.

Abstract: Situated in today's digital context, where the open science movement is a new paradigm for the production and circulation of knowledge, this paper analyzes the extent to which research data repositories meet the technical, legal, and ethical needs of researchers in the education and the social sciences domains. The methodological approach combines a literature review, an expert interview, and analysis of repository catalogs. From the fusion of the obtained results, it can be concluded that the characteristics of repositories that are most valued by researchers in these disciplines are reputation, ease of use, availability of contextual information, sustainability (long-term data preservation), technical support, and compliance with FAIR principles (which promote the production of Findable, Accessible, Interoperable, and Reusable data). In addition, this article delves into the role that repository catalogs play in facilitating the selection of appropriate repositories for depositing data, providing researchers with useful information for determining whether specific repositories meet their needs. This study provides valuable information for promoting the adoption of open data practices in the field of educational research.

Keywords: Research Tools, Archives, Online Catalogs, Metadata, Data Use.

1. Introducción

En la era digital actual, el movimiento de ciencia abierta ha reconfigurado profundamente los paradigmas de producción y circulación del conocimiento científico. Dentro de este marco transformador, el acceso público a los datos de investigación emerge como un componente estratégico fundamental, particularmente en el campo educativo. La aplicación de estos paradigmas no solo robustece los pilares de transparencia y reproducibilidad que sustentan el método científico (OECD 2020), sino que además facilita el análisis secundario y contribuye al impacto social del conocimiento (European Commission 2016; Wilkinson et al. 2016). Los beneficios potenciales son particularmente significativos en el ámbito educativo, donde la posibilidad de contrastar resultados entre diferentes contextos socioculturales y sistemas pedagógicos podría generar insumos para la evolución de las políticas públicas en educación.

La UNESCO (2021, p. 10) conceptualiza los datos abiertos de investigación como aquellos que deben estar «disponibles de manera oportuna, en un formato fácil de utilizar, legible y modificable por personas y máquinas», articulándose bajo los principios FAIR (Fáciles de encontrar, Accesibles, Interoperables y Reutilizables) (Wilkinson et al. 2016). La disponibilidad de repositorios alineados con estos principios facilita la tarea de los investigadores, tanto para depositar como para reusar datos. Sin embargo, no es trivial decidir en qué repositorio depositar, organizar y gestionar los datos de investigación. En este sentido, los principios TRUST (Lin et al. 2020) representan un marco de referencia muy valioso para evaluar la calidad de los repositorios científicos, promoviendo la Transparencia (políticas de uso claras), Responsabilidad (en lo que refiere a la gestión de las colecciones de datos), el foco en el Usuario (disponibilidad de herramientas para explorar, ubicar y reusar datos), Sostenibilidad (disponibilidad de los servicios de acceso a los datos a largo plazo) y Tecnología (uso e implementación de estándares y herramientas adecuadas para la gestión y curación de los datos). La exhaustiva caracterización de Behnke et al. (2020) detalla cómo estos principios incluyen tanto requerimientos organizacionales (como políticas claras de gestión de datos y soporte para formatos estándar) como técnicos (desde metadatos granularizados hasta sistemas de identificadores persistentes).

La organización y agrupación de colecciones de datos de investigación en repositorios responde a diferentes criterios de acuerdo a los cuales es posible categorizar los repositorios como disciplinares, cuando contienen datos de un área específica de conocimiento, generalistas, en caso contrario, y también institucionales, en caso de albergar datos de una institución en particular. A su vez, en algunas disciplinas, por ejemplo economía o epidemiología, la mayor parte de los datos gestionados son cuantitativos, mientras que en otras, como las ciencias sociales y la educación, los datos son en su mayor parte cualitativos. Cada una de estas categorías de repositorios puede requerir la priorización de algunos principios sobre otros, o el uso de diferentes estándares o tecnologías para su implementación. En general, los datos de investigación de una misma disciplina tienen características en común, por lo que los requerimientos tecnológicos son similares. En particular, la investigación educativa enfrenta obstáculos específicos, especialmente cuando se trabaja con datos cualitativos que, como destaca Antonio et al. (2019), demandan niveles excepcionales de contextualización y mecanismos especializados de protección ética, implementados con tecnologías específicas, que muchos repositorios generalistas no están preparados

para ofrecer. Otros estudios como el de Strecker et al. (2023) informan sobre los riesgos de estos repositorios para cumplir con los estándares de calidad, indicando además que los repositorios disciplinares afines a humanidades y ciencias sociales son de los más afectados por este tipo de problemas. Otros autores (Benjelloun et al., 2020; Gerasimov et al. 2024) también advierten sobre tendencias de centralización de la mayoría de los datasets sobre el conjunto de repositorios generalistas. Paradójicamente, varias recomendaciones de estos repositorios generalistas alientan a los investigadores a utilizar repositorios disciplinares e institucionales (Barbosa et al. 2024). En particular, los trabajos de Song et al. (2025) y Ávila Barrientos (2024) destacan por su análisis exhaustivo de las pautas para seleccionar repositorios de datos científicos confiables. También organismos internacionales como la UNESCO (2021) y la OECD (2020), alineados con los principios FAIR y TRUST, han impulsado políticas que fomentan el uso de repositorios digitales especializados para el almacenamiento, preservación y diseminación ética de datos científicos.

Como argumentan Borgman et al. (2018), estas infraestructuras digitales han dejado de ser meros almacenes de información para convertirse en nodos fundamentales de las redes contemporáneas de producción de conocimiento, facilitando no solo el acceso sino también la interoperabilidad y el enriquecimiento progresivo de los datasets mediante sucesivas reutilizaciones. En esta misma línea, la reciente conceptualización de (Ávila Barrientos 2024, p. 27) enfatiza su rol como sistemas dinámicos que permiten «almacenar, organizar, recuperar y acceder a conjuntos de datos de diversa naturaleza temática y tipológica», constituyéndose así en motores esenciales para la ciencia abierta.

La aplicación de los principios FAIR (Wilkinson et al. 2016) y TRUST (Lin et al. 2020) enfrentan retos particulares en disciplinas como la educación y las ciencias sociales, donde la heterogeneidad, sensibilidad y complejidad de los datos dificultan la estandarización y la creación de infraestructuras especializadas, alineadas con los requerimientos de estas disciplinas. Sin embargo, la escasez de repositorios disciplinares en estas áreas, sumada a desafíos éticos, legales y culturales, obliga a los investigadores a recurrir mayoritariamente a repositorios institucionales o generalistas, que no siempre satisfacen los requerimientos específicos para la gestión óptima de datos educativos y sociales.

Como advierte Borgman (2016) en su análisis sobre las ecologías del conocimiento, los datos solo alcanzan su pleno valor científico cuando se integran en ecosistemas robustos que articulen coherentemente dimensiones técnicas, prácticas comunitarias e institucionales. Precisamente, esta articulación presenta desafíos particulares en el ámbito educativo y de ciencias sociales, donde la disponibilidad de repositorios disciplinares especializados es escasa en comparación con otras áreas del conocimiento (Krähmer et al., 2023; Lamb et al., 2024). Esta disparidad resulta paradójica si consideramos la complejidad intrínseca de los datos educativos, que abarcan desde mediciones cuantitativas estandarizadas hasta registros etnográficos profundamente contextualizados, frecuentemente conteniendo información sensible (por ej. sobre menores de edad en entornos escolares específicos (Gomes et al. 2022).

Frente a estos desafíos, iniciativas como *re3data* (Pampel et al., 2013) han buscado crear catálogos integrados que faciliten la localización y evaluación de repositorios especializados. La selección informada de repositorios se vuelve entonces un desafío crucial. Los catálogos de repositorios emergen como herramientas clave

para buscar, filtrar y evaluar opciones según criterios técnicos, éticos y disciplinares. Sin embargo, en lo que refiere a repositorios de datos de investigaciones en educación, se observa una muy escasa presencia disciplinar. A modo de ejemplo, de acuerdo a la información recabada en junio de 2025, los catálogos *FAIRSharing*, *DataCite Commons* y *re3Data* tienen registrados en total 1252, 1767 y 3259 repositorios respectivamente, de los cuales únicamente 22, 5 y 47 corresponden a repositorios de datos de investigación en el área de la educación. En este sentido, creemos que este trabajo aporta en prestar atención a las necesidades de los investigadores de la educación, para continuar mejorando el servicio que los repositorios pueden ofrecerles, para incentivar así su desarrollo.

Este artículo analiza las necesidades para la publicación de datos en educación y ciencias sociales, y las características de los repositorios más valorados por las investigadoras y los investigadores de estas disciplinas, considerando sus preocupaciones éticas, técnicas y legales. Además, este trabajo explora el papel de los catálogos de repositorios en la selección de repositorios como toma de decisiones informadas, con el objetivo de fortalecer la cultura de datos abiertos y contribuir al avance de la ciencia abierta en estas disciplinas.

2. Metodología

Para la realización de este estudio se siguió un diseño metodológico de enfoque secuencial exploratorio con las siguientes tres fases interrelacionadas: Revisión Bibliográfica, Entrevista a Experto y Análisis de Catálogos. Su objetivo fue identificar cómo responden los repositorios a las preocupaciones éticas, técnicas y legales que tienen las investigadoras y los investigadores en el momento de depositar datos de investigaciones educativas.

2.1. Fases Metodológicas

- Fase 1. Revisión Bibliográfica. Se utilizaron diversas fuentes: artículos en Scopus/WoS/ (keywords: "open data" + "education research" + "concerns"), informes de UNESCO y OECD. Criterios de exclusión: trabajos sobre repositorios de recursos educativos abiertos y trabajos que solo consideran el acceso abierto a publicaciones de investigación. Para el análisis se aplicó una codificación asistida con Elicit¹ para categorizar preocupaciones recurrentes (ej. privacidad, propiedad intelectual, esfuerzo adicional). Y el resultado esperado, un mapa conceptual de barreras documentadas en la literatura (2015–2025).
- Fase 2. Entrevista a Experto. Se entrevistó al responsable del repositorio de datos abiertos de la Agencia Nacional de Investigación e Innovación (ANII) de Uruguay. El instrumento (protocolo) fue una entrevista semiestructurada, con una duración de 45 minutos. Para el análisis se llevó a cabo una transcripción y codificación con el fin de rectificar o ratificar categorías revisadas en la literatura.
- Fase 3. Análisis de Catálogos de Repositorios. Se analizó una muestra compuesta por 6 Catálogos de amplio uso internacional (OpenDOAR, ROAR, OpenAIRE, FAIRSharing, DataCite Commons). Las variables evaluadas fueron las siguientes: actividad del repositorio, creación de identificadores persistentes, soporte para especificar el

¹ Asistente de investigación con IA: <https://elicit.com/>

contexto, cumplimiento de estándares para interoperabilidad, adhesión a políticas de licenciamiento y sostenibilidad.

Para la triangulación y validación de datos se combinaron los tres métodos, revisión bibliográfica, consulta a experto y análisis de catálogos, lo que permitió triangular la información, aumentando la validez y confiabilidad de los resultados.

3. Resultados

Esta sección presenta los resultados obtenidos a partir de la revisión bibliográfica, la consulta a expertos y el análisis de catálogos de repositorios.

3.1. Resultados de revisión bibliográfica

Numerosos estudios se han centrado en identificar los factores que impactan en la voluntad de los investigadores en compartir sus datos en abierto y se ha hecho un esfuerzo por destacar algunos de ellos aquí, especialmente desde la perspectiva de los repositorios. La revisión bibliográfica realizada muestra que los estudios abarcan encuestas, revisiones sistemáticas de la literatura, meta-análisis, grupos focales y mesas redondas, de las más actuales (Tenopir et al., 2011; Zuidervijk et al., 2020; Barczak et al., 2022; Casali et al., 2022; Thøgersen & Borlund, 2022; Borycz et al., 2023; Logan et al., 2021; Mosha, 2024; Shamsudin et al., 2025; Grattarola et al., 2024). Es interesante observar que los estudios son con investigadores de distintas disciplinas, distinto nivel de *expertise* y de diferentes contextos geográficos. Curiosamente, pocos de estos estudios se han realizado con investigadores de Ciencias de la Educación. Entre los más destacados que involucran opiniones de investigadores relacionados con la disciplina Educación encontramos el trabajo de Casali, Motz, and Sprock (2022) que analiza una encuesta administrada a investigadores de la comunidad LACLO (Conferencia Latinoamericana de Tecnologías de Aprendizaje) y los trabajos de Logan, Hart, and Schatschneider (2021) y de Neild et al. (2022) que presentan guías sobre publicación de datos abiertos expresamente dirigidas para investigadores en Educación.

Mientras que el trabajo de Casali et al. (2022) destaca barreras como el tiempo y el esfuerzo requeridos para compartir los datos, la falta de financiamiento para la estandarización de los datos y las restricciones relacionadas con la seguridad y la confidencialidad de los datos, observamos que en los trabajos de Logan et al. (2021) y Neild et al. (2022) se presta especial atención a los procedimientos para mantener la seguridad y la confidencialidad de los datos, evidenciando que éste constituye un problema en común.

La mayoría de los trabajos coinciden en reconocer como aspectos importantes en la problemática planteada por los investigadores el esfuerzo extra que significa curar los datos para su publicación en abierto. Mientras en general se atribuye esta problemática al desarrollo de los repositorios poco centrados en el usuario, para otros autores como Borycz et al. (2023) y Casali et al. (2022), se debe a la falta de apoyo administrativo a los investigadores en la preparación de los datos.

Otros estudios, como el de Grattarola et al. (2024) y Mabile et al. (2025), señalan que la desmotivación principal para compartir los datos en abierto recae en la falta de incentivos institucionales y la falta de reconocimiento por la publicación de datos.

Respecto al caso específico de datos procedentes de investigaciones cualitativas, se destaca en todos los trabajos la problemática de contar con elementos que permitan especificar el contexto en el cual los datos fueron obtenidos y la preocupación por contar con curaduría específica para asegurar el resguardo del investigador ante el problema legal y ético de uso de datos sensibles.

Los investigadores cualitativos en el área de ciencias sociales mencionan temas relativos a la privacidad de sus datos y el compromiso ético que implica su publicación. En ciencias sociales, los datos suelen involucrar personas, y aunque se anonimicen, muchos investigadores temen que los participantes puedan ser reidentificados, especialmente en comunidades pequeñas o temas sensibles (Krähmer et al., 2023). A pesar de existir mecanismos para mitigarlo, como el consentimiento informado, la responsabilidad de los repositorios en cuanto a la custodia de los datos no siempre resulta fácil de comprender para los productores de datos (Prosser et al., 2022).

Estudios recientes como el de Lamb et al. (2024) ratifican algunos desafíos propios de las investigaciones cualitativas. Los investigadores cualitativos destacan que los datos (como entrevistas o diarios de campo) necesitan un conocimiento contextual profundo para ser interpretados adecuadamente cuando otros investigadores deciden reusar el dataset.

3.2. Resultado consulta a experto

Las conclusiones derivadas de la entrevista con el experto se basan en un estudio interno realizado por la Agencia Nacional de Investigación e Innovación (ANII). Este estudio, aunque no publicado, proporcionó información valiosa sobre las perspectivas de los investigadores en el contexto de Uruguay. El estudio de la ANII reveló que los investigadores tienden a percibirse principalmente como consumidores de datos, más que como productores. Esta percepción influye en sus expectativas y necesidades en relación con los repositorios de datos. El experto entrevistado, basándose en los resultados del estudio, destacó tres características clave que los investigadores consideran esenciales al momento de publicar sus datos:

- a) Asignación de identificadores persistentes: La asignación de identificadores únicos y permanentes, preferentemente DOI, es vista como un requisito fundamental, especialmente debido a que muchas revistas científicas los exigen para la evaluación y citación de artículos.
- b) Sostenibilidad y perdurabilidad: La capacidad del repositorio para garantizar la conservación de los datos a largo plazo es considerada crucial para asegurar su disponibilidad y reutilización futura.
- c) Interoperabilidad: La capacidad del repositorio para facilitar la integración y el intercambio de datos entre diferentes plataformas y sistemas es valorada como esencial para maximizar el impacto y la reutilización de los datos.

La modalidad semiestructurada de la entrevista facilitó la exploración profunda de temas relevantes, sin restringir la espontaneidad del diálogo, lo que contribuyó a validar y ampliar los resultados del relevamiento bibliográfico con evidencia empírica y experiencia directa.

3.3. Resultado del análisis de catálogos de repositorios

Los estudios de Witt et al. (2024) y Song et al. (2025) establecen los principales criterios para evaluar repositorios de datos de investigación; no obstante, señalan que los investigadores enfrentan obstáculos para verificar el cumplimiento de dichos criterios, dada la escasa accesibilidad de esta información a priori. Esta limitación dificulta la selección de un repositorio adecuado para la publicación y la reutilización de datos. En este trabajo, analizamos el papel de los catálogos de repositorios en la difusión eficiente de dicha información, tomando como referencia los criterios definidos por estos autores y los desarrollados en el *Data Repository Attributes Working Group*². Las características que buscamos identificar son; actividad del repositorio, creación de identificadores persistentes, soporte para especificar el contexto, cumplimiento de estándares para interoperabilidad y adhesión a políticas de licenciamiento y sostenibilidad.

Actividad del repositorio

Los factores de actividad del repositorio y especificidad disciplinar determinan fuertemente el impacto en la visibilidad y la reutilización de los datos. La actividad reciente del repositorio es un indicador clave de su confiabilidad. Son indicadores de actividad la frecuencia de actualizaciones (ej., nuevos datasets depositados mensualmente), fecha del último dataset depositado (evitar repositorios sin actividad en los últimos 12 meses), estadísticas públicas (ej., número de descargas o citas en *DataCite*). Por otro lado, un repositorio especializado en un área temática ofrece la ventaja significativa de utilizar esquemas de metadatos específicos (ej. DDI³ para ciencias sociales), lo que mejora la interoperabilidad y el descubrimiento de los datos. Además, los repositorios temáticos atraen a usuarios especializados, lo que incrementa el impacto y la citación del dataset.

Creación de identificadores persistentes

El Identificador Digital de Objeto (DOI, por sus siglas en inglés) desempeña un papel fundamental en los repositorios de datos de investigación, ya que garantiza la localización, citación y preservación permanente de los conjuntos de datos. Proporciona un enlace único y estable, evitando la pérdida de datos debido a cambios en las URL o la desactualización de repositorios. Esto asegura que los conjuntos de datos permanezcan accesibles a largo plazo, cumpliendo con los principios FAIR. Al integrarse con sistemas de gestión bibliográfica (como *Crossref* o *DataCite*), el DOI permite vincular conjuntos de datos con artículos científicos, mejorando su visibilidad y facilitando el seguimiento de su impacto.

El DOI (*Digital Object Identifier*) es asignado por agencias registradoras autorizadas, siendo *DataCite* la más utilizada para datos de investigación o por repositorios/instituciones afiliadas a *DataCite*. En algunos casos puede ocurrir que un mismo dataset termine teniendo múltiples DOIs. Esto puede ocurrir si un investigador carga el mismo dataset en distintos repositorios y cada uno le asigna un DOI distinto, o en caso de actualizaciones o versiones, ya que algunos repositorios generan un nuevo DOI para versiones corregidas o ampliadas del mismo dataset. Esto no es deseable ya que las citas al dataset se dispersan entre varios DOIs, subestimando su impacto real y además se dificulta identificar cuál es la versión más actualizada. Para evitar estos problemas se recomienda priorizar el uso de repositorios de referencia en cada

² <https://www.rd-alliance.org/groups/data-repository-attributes-wg/>

³ <https://ddialliance.org/>

disciplina (ej. *ICPSR* en ciencias sociales) para evitar duplicaciones. En caso de esto no ser posible, las soluciones recomendadas son usar repositorios que admitan versiones bajo un mismo DOI y vincular DOIs alternativos mediante metadatos (relaciones *isVersionOf* o *isIdenticalTo* en esquemas como Schema.org).

Soporte para documentar el contexto

El contexto de un dataset incluye detalles sobre su origen, metodología con la cual fue generado, uso potencial y limitaciones. Los repositorios recolectan la información de contexto a través de metadatos descriptivos, documentación técnica (como Protocolos de laboratorio, cuestionarios o scripts de procesamiento), documentos adicionales (como Readme, Guías e Informes) y publicaciones vinculadas (DOI de artículos relacionados que usaron/citaron el dataset). En los metadatos descriptivos al menos se debe informar:

- a) Origen: Quién lo creó (institución, proyecto de investigación, autoría) y Cobertura temporal/geográfica (cuándo y dónde se generó el dataset).
- b) Propósito: Objetivo de la recolección.
- c) Métodos de recolección: Instrumentos, protocolos o software utilizados.
- d) Licencias y restricciones: Condiciones de uso (CC-BY, GDPR, acceso restringido).

Los catálogos como *OpenAIRE* y *DataCite Commons* fomentan el uso de estándares internacionales (como el protocolo OAI-PMH y metadatos FAIR), lo que asegura que los datos sean fácilmente localizables, accesibles y reutilizables por diferentes plataformas y sistemas, favoreciendo la colaboración y el intercambio de información entre comunidades científicas. Esta buena práctica de usar vocabularios controlados y ontologías disciplinares (p. ej., *OECD Education Taxonomy*) para facilitar la interoperabilidad es impulsada por FAIRsharing ofreciendo entre sus funcionalidades un directorio curado de estándares de metadatos. Cuando un repositorio se indexa en *FAIRsharing*, éste evalúa y recomienda estándares de metadatos específicos según su disciplina.

Adhesión a políticas

Otro de los aspectos a considerar de un repositorio, que dependen de factores políticos o legales, son sus políticas de uso y de preservación o sostenibilidad. El problema de repositorios que se extinguen es un problema grave como reportan Strecker et al. (2023). Algunos catálogos, como *FAIRsharing*, facilitan la evaluación de estos factores, permitiendo identificar la relación entre las políticas específicas de cada repositorio con políticas más generales. Si tomamos el ejemplo de la NIH⁴ como financiador, este podría imponer el cumplimiento de sus términos de uso de datos⁵. *FAIRsharing* es capaz de desplegar los repositorios compatibles con esta política, visualizados mediante el grafo de la figura 1. De esta forma, un investigador con un proyecto financiado por *NIH* puede limitar la selección del repositorio a utilizar por los encontrados en la consulta que produce el grafo de la figura 1.

⁴ <https://www.nih.gov/>

⁵ [nih_policy_fairsharing](https://fairsharing.org/fairsharing/7861ef)] 2023 Final NIH Policy for Data Management and Sharing - <https://fairsharing.org/fairsharing/7861ef>

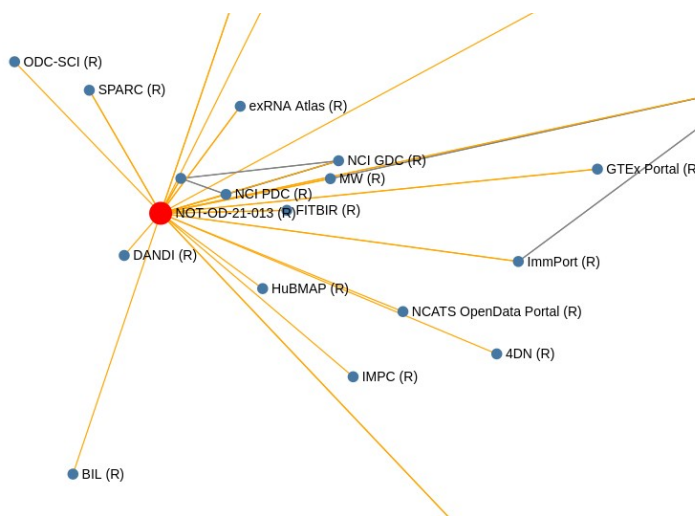


Figura 1. Grafo de repositorios que utilizan la política del NIH - FAIRSharing (Obtenido el 26/06/2025).

Otro mecanismo que aporta en la identificación de las adhesiones de los repositorios a ciertas políticas de uso y de sostenibilidad son las Certificaciones. Las certificaciones de repositorios son mecanismos de evaluación externa que garantizan que un repositorio cumple con estándares de calidad, preservación y buenas prácticas en la gestión de datos. Las certificaciones son especialmente útiles para visualizar si los repositorios cumplen con políticas de preservación o sostenibilidad. Por ejemplo, un repositorio con alguna certificación indica que debió demostrar alguna de estas funcionalidades: asigna DOIs, usa esquemas de metadatos disciplinares, cuenta con una infraestructura técnica sostenible (backups, migración de formatos), ofrece gestión de versiones para el control de cambios en los datasets, tiene un soporte técnico para dar respuesta a incidencias y que está respaldado por una institución estable (universidad, organismo público). Distintos sellos de certificaciones evalúan diferentes funcionalidades. Por ejemplo, un repositorio con certificación «*CoreTrustSeal*»⁶ está verificado para los criterios de asignación de DOI y preservación por un período mayor a 10 años, mientras que con un sello ISO 16363⁷ tiene una verificación exhaustiva en los criterios de cumplimiento legal (ej., GDPR, propiedad intelectual) pero no en asignación de DOI.

Los catálogos recopilan datos de organismos certificadores y estandarizan la visualización de certificaciones para facilitar la comparación. Por ejemplo, *re3data* muestra el año de certificación y su vigencia. *FAIRsharing* no certifica, pero evalúa y recomienda repositorios que adhieren al uso de estándares de metadatos y a principios FAIR. Además incluye un filtro por certificación (ej., *CoreTrustSeal*, ISO 16363⁷) y proporciona enlaces a los informes de certificación oficiales. Un investigador que busca un repositorio para datos de educación puede usar *FAIRsharing* y filtrar por *CoreTrustSeal* y verificar en *re3data* si el repositorio tiene certificación vigente.

Compatibilidad de los repositorios con servicios

⁶ <https://www.coretrustseal.org>

⁷ <https://www.iso.org/standard/87472.html>

Conocer la capacidad del repositorio para proveer o integrarse con ciertos servicios especializados en el área de investigación, así como las condiciones bajo las cuales los provee (precio del servicio o el nivel de disponibilidad del servicio), puede resultar relevante para los investigadores. Para enumerar algunos de estos servicios y funcionalidades se pueden mencionar la capacidad de realizar o recibir revisiones formales, la posibilidad de establecer períodos de embargo, asistencia técnica especializada o la disponibilidad de servicios de curaduría disciplinar. En particular, para el caso de repositorios de datos de investigación cualitativos, la presencia o integración de servicios como anonimización, o mecanismos de control de acceso resulta valiosa para los investigadores. Catálogos como *OpenAIRE* ofrecen información sobre la integración de los diferentes repositorios con el ecosistema de servicios de la *EOSC* por ejemplo con el filtro «*Compatibility Level*» se pueden visualizar el nivel de compatibilidad del repositorio con servicios como: «*Amnesia*»⁸ servicio de anonimización, *OpenAIRE AAI*⁹, API para servicio de acceso, o «*Argos*»¹⁰ servicio de Plan de Gestión de Datos, entre otros. La tabla 1 ofrece un resumen de las principales características de los catálogos de repositorios revisados.

Tabla 1. Principales características de los catálogos de repositorios.

Catálogo	Características
OpenDOAR	Permite conocer información básica del repositorio.
ROAR	Similar a OpenDOAR, pero permite conocer la disciplina asociada al repositorio. Especializado en repositorios de publicaciones académicas, contiene repositorios de datos de investigación de manera casual.
Re3data	Se especializa en describir repositorios de datos de investigación. Provee la mayor cantidad de registros de repositorios. Muy útil su publicación de vigencia de certificaciones.
FAIRsharing	Posee la capacidad de navegar entre las relaciones de los estándares y políticas con los diferentes repositorios de datos. Dentro de los estándares incluye los principios FAIR, TRUST y distintas certificaciones. Incluye repositorios variados (de datos, documentos, multimedia, software, etc.).
DataCite Commons	Integra información sobre repositorios de investigación de diversa índole, y destaca por proveer información relativa al contenido de los mismos de manera sintética, aunque no permite conocer muy a detalle algunas características propias del repositorio ni proporciona criterios de filtrado avanzados como Re3data.
OpenAIRE	Similar a DataCite, permite conocer, aunque no navegar entre las relaciones de los repositorios con tanta facilidad como FAIRSharing. Destaca la calidad con la que se maneja la procedencia de los datos, permite conocer que repositorios comparten contenidos y de cual catálogo proviene la información del repositorio.

4. Conclusión

A continuación, se presentan los hallazgos obtenidos a partir de los resultados de la aplicación de la metodología. Los hallazgos se sintetizan en la tabla 2. En la primera

⁸ <https://www.openaire.eu/amnesia-guide>

⁹ <https://catalogue.openaire.eu/service/openaire.aai/overview>

¹⁰ <https://catalogue.openaire.eu/service/openaire.argos/overview>

columna se muestran las características de los repositorios que cumplen con las necesidades planteadas por los investigadores listados en la segunda columna, y en la tercera columna se listan los catálogos que, en menor o mayor medida, proveen información que permite a los investigadores conocer qué repositorios cumplen con la característica.

Tabla 2. Características de repositorios más valoradas por investigadores.

Característica	Necesidad del investigador	Catálogo que provee la información
Actividad, Reputación	Reconocimiento	Re3data, FAIRSharing, OpenAIRE
Apoyo técnico	Menos esfuerzo	FAIRSharing, DataCite Commons
Disponibilidad de metadatos	Información de contexto	FAIRSharing, DataCite Commons, OpenAIRE
Cumplimiento de principios FAIR	Acceso y reuso de datos, interoperabilidad	Re3data, FAIRSharing, OpenAIRE
Sostenibilidad, Adhesión a políticas	Disponibilidad e integridad de datos	Re3data, FAIRSharing

Las características presentadas en la primera columna de la tabla 2 representan los atributos de los repositorios que se identifican como los más valorados por los investigadores, en el presente trabajo. Son los siguientes: (a) Actividad y Reputación: El repositorio debe cumplir con estándares de la comunidad disciplinar, tener actividad reconocida gozando de buena reputación dentro de la comunidad científica y contar con el respaldo de organizaciones profesionales o instituciones académicas. (b) Apoyo técnico: El repositorio debe proporcionar documentación completa y un buen soporte técnico para facilitar su uso. (c) Disponibilidad de metadatos: El repositorio debe contener metadatos descriptivos que proporcionen información de contexto de los datos, como el autor y propósito de la recolección, y la metodología utilizada. (d) Cumplimiento de los principios FAIR: El repositorio debe proporcionar metadatos claros y admitir la interoperabilidad entre diferentes formatos de datos. Cumplir con una estructura de gobierno clara y políticas transparentes que permitan la reusabilidad segura y ética. (e) Sostenibilidad y adhesión a políticas: El repositorio debe contar con un modelo de financiación a largo plazo y un plan para garantizar su funcionamiento continuo y contempla planes de continuidad para evitar pérdida de información. El repositorio debe contar con una infraestructura sólida en términos de capacidad de almacenamiento, seguridad de los datos y protocolos de respaldo.

4.1. Discusión

Es importante destacar la apertura de los investigadores hacia la publicación de datos y su reconocimiento de los beneficios que esto conlleva tanto para quienes publican como para quienes acceden a los datos. Sin embargo, los repositorios de datos actuales no parecen ofrecer una solución integral que cubra todas las necesidades de los investigadores educativos.

A pesar de esta limitación, el estudio permitió realizar una evaluación comparativa de los catálogos de repositorios de datos de investigación. En este

sentido, *FAIRSharing* se destaca como la herramienta que ofrece los mejores niveles de curaduría y calidad de datos. También se resalta su modelo de sostenibilidad y la transparencia de sus criterios y procesos. *FAIRSharing* proporciona una visión orgánica del proceso de publicación y de la infraestructura de investigación abierta. Sin embargo, no se recomienda su uso en exclusividad. Debido a que la base de datos de *FAIRSharing* carece de algunos registros, se recomienda realizar una consulta paralela en *DataCite Commons* y para obtener una mejor visualización de las vigencias de las certificaciones se recomienda usar *re3data*.

Frente a las limitaciones que perjudican la visibilidad, varios autores concuerdan en que es necesario mejorar la capacidad de descubrimiento de los repositorios y servicios disponibles, como lo señala recientemente Huber et al. (2024). En ese mismo trabajo, también se señala la necesidad de mejorar la capacidad de los repositorios y la infraestructura para capturar diversa información disciplinar. Trabajos como Ulrich et al. (2024) proponen mecanismos para descubrir los servicios y capacidades de cada repositorio, aunque siguen requiriendo esfuerzos de estandarización por parte de los gestores de repositorios para ser de utilidad. Este tipo de propuestas potenciaría enormemente la calidad de los datos recolectados por los catálogos de repositorios, haciéndolos mucho más confiables y descriptivos.

Otros artículos como Vogt et al. (2025) incluso desdibujan el rol de los repositorios y se centran en los contenidos, proponiendo crear una capa de interoperabilidad semántica que logre unificar los recursos desde una perspectiva conciliadora para cada disciplina, facilitando el acceso por parte de los investigadores no solo a los datos de investigación, sino a los recursos en general, desdibujando la línea que diferencia a los artículos de investigación de los datos de investigación o incluso de recursos externos al área de la investigación (eg: REAs, datos gubernamentales, software, etc.), conceptos que son difíciles de diferenciar para el grueso de los investigadores y que muchas veces puede no tener sentido diferenciar desde el punto de vista del investigador.

Otras estrategias sugieren el uso de planes de gestión de datos (PGD) propuestos por Bicarregui et al. (2012) para transparentar el diálogo entre los diferentes actores durante el proceso de investigación y, en un paso posterior, plasmarlo en un documento estructurado y accionable automáticamente (maPGD), como propone Miksa (2019). Si bien el uso de PGDs no es nuevo, aún requiere incrementar su adopción antes de ser explotado en su totalidad.

Este tipo de iniciativas contribuirán a acortar la brecha disciplinaria y promover el diálogo entre la gestión de datos y el conocimiento disciplinar.

5. Referencias

- Antonio, M. G., Schick-Makaroff, K., Doiron, J. M., Shields, L., White, L., & Molzahn, A. (2020). Qualitative Data Management and Analysis within a Data Repository. *Western Journal of Nursing Research*, 42(8), 640-648.
<https://doi.org/10.1177/0193945919881706>
- Ávila Barrientos, E. (2024). Explorando el papel de los repositorios de datos de investigación en el contexto de la ciencia abierta. *Métodos de información*, 15(28), 1-29. <https://doi.org/10.5557/IIMEI15-N28-001029>
- Barbosa, S., Carson, M., Curtin, L., Diggs, S., Gonzales, S., Herzog, J., Olson, E., Snowden, T., Sweet, S., Van Gulick, A., & Wood, J. (2024). *Generalist Repository Selection Flowchart* [Graphic]. Zenodo.
<https://doi.org/10.5281/ZENODO.11105430>
- Barczak, G., Hopp, C., Kaminski, J., Piller, F., & Pruschak, G. (2022). How open is innovation research? – An empirical analysis of data sharing among innovation scholars. *Industry and Innovation*, 29(2), 186-218.
<https://doi.org/10.1080/13662716.2021.1967727>
- Behnke, C., Bonino, L., Coen, G., Le Franc, Y., Parland-von Essen, J., Riungu-Kalliosaari, L., & Staiger, C. (2020). *D2.3 Set of FAIR data repositories features*.
<https://doi.org/10.5281/ZENODO.5361952>
- Benjelloun, O., Chen, S., & Noy, N. (2020). Google Dataset Search by the Numbers. En J. Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, & L. Kagal (Eds.), *The Semantic Web – ISWC 2020* (Vol. 12507, pp. 667-682). Springer International Publishing.
https://doi.org/10.1007/978-3-030-62466-8_41
- Bicarregui, J., Gray, N., Henderson, R., Jones, R., Lambert, S., & Matthews, B. (2012). *DMP Planning for Big Science Projects* (No. arXiv:1208.3754). arXiv.
<https://doi.org/10.48550/arXiv.1208.3754>
- Borgman, C. L. (2016). *Big data, little data, no data: Scholarship in the networked world*. The MIT Press.
- Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, 70(8), 888-904.
<https://doi.org/10.1002/asi.24172>
- Borycz, J., Olendorf, R., Specht, A., Grant, B., Crowston, K., Tenopir, C., Allard, S., Rice, N. M., Hu, R., & Sandusky, R. J. (2023). Perceived benefits of open data are improving but scientists still lack resources, skills, and rewards. *Humanities and Social Sciences Communications*, 10(1), 339. <https://doi.org/10.1057/s41599-023-01831-7>
- Casali, A., Motz, R., & Sprock, A. S. (2022). Open Data Management of Research in Education: Knowledge and practices in LATAM. *2022 XVII Latin American Conference on Learning Technologies (LACLO)*, 1-7.
<https://doi.org/10.1109/LACLO56648.2022.10013458>
- European Commission. (2016). *Guidelines on FAIR data management in H2020*. http://archive.org/details/h2020-hi-oa-data-mgt_en
- Gerasimov, I., Kc, B., Mehrabian, A., Acker, J., & McGuire, M. P. (2024). Comparison of datasets citation coverage in Google Scholar, Web of Science, Scopus, Crossref, and DataCite. *Scientometrics*, 129(7), 3681-3704. <https://doi.org/10.1007/s11192-024-05073-5>
- Gomes, D. G. E., Pottier, P., Crystal-Ornelas, R., Hudgins, E. J., Foroughirad, V., Sánchez-Reyes, L. L., Turba, R., Martínez, P. A., Moreau, D., Bertram, M. G., Smout, C. A., & Gaynor, K. M. (2022). Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences*, 289(1987), 20221113.
<https://doi.org/10.1098/rspb.2022.1113>
- Grattarola, F., Laufer, G., Rodríguez-Tricot, L., González, E. M., & Mello, F. T. de. (2024).

- Desafíos y barreras para la apertura de datos de biodiversidad en Uruguay. *Ecología Austral*, 34(3), Article 3. <https://doi.org/10.25260/EA.24.34.3.0.2413>
- Grattarola, F., Shmagun, H., Erdmann, C., Cambon-Thomsen, A., Thomsen, M., Kim, J., & Mabile, L. (2024). Gaps between Open Science activities and actual recognition systems: Insights from an international survey. *PLOS ONE*, 19(12), e0315632. <https://doi.org/10.1371/journal.pone.0315632>
- Huber, R., Gonzalez Beltran, A., Neidiger, C., Ulrich, R., & L'Hours, H. (2025, enero 20). *Towards Transparent Presentation of FAIR-enabling Data Repository Functions & Characteristics*. <https://doi.org/10.5194/egusphere-egu24-20387>
- Krähmer, D., Schächtele, L., & Schneck, A. (2023). Care to share? Experimental evidence on code sharing behavior in the social sciences. *PLOS ONE*, 18(8), e0289380. <https://doi.org/10.1371/journal.pone.0289380>
- Lamb, D., Russell, A., Morant, N., & Stevenson, F. (2024). The challenges of open data sharing for qualitative researchers. *Journal of Health Psychology*, 29(7), 659-664. <https://doi.org/10.1177/13591053241237620>
- Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giarretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, 7(1), 144. <https://doi.org/10.1038/s41597-020-0486-7>
- Logan, J. A. R., Hart, S. A., & Schatschneider, C. (2021). Data Sharing in Education Science. *AERA Open*, 7, 23328584211006475. <https://doi.org/10.1177/23328584211006475>
- Mabile, L., Shmagun, H., Erdmann, C., Cambon-Thomsen, A., Thomsen, M., & Grattarola, F. (2025). Recommendations on Open Science Rewards and Incentives: Guidance for Multiple Stakeholders in Research. *Data Science Journal*, 24. <https://doi.org/10.5334/dsj-2025-015>
- Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLOS Computational Biology*, 15(3), e1006750. <https://doi.org/10.1371/journal.pcbi.1006750>
- Mosha, N. F. V. (2024). Open Science Movement and Researchers' Perspectives in Higher Learning Institutions: A Conceptual Framework. *Pakistan Journal of Life and Social Sciences (PJLSS)*, 22(2). <https://doi.org/10.57239/PJLSS-2024-22.2.00998>
- Neild, R. C., Robinson, D., & Agufa, J. (2022). *Sharing Study Data: A Guide for Education Researchers*. (NCEE 2022-004). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. https://ies.ed.gov/sites/default/files/migrated/nces_pubs/ncee/pubs/2022004/pdf/2022004.pdf
- OECD. (2020). *Enhanced Access to Publicly Funded Data for Science, Technology and Innovation*. OECD. <https://doi.org/10.1787/947717bc-en>
- Pampel, H., Vierkant, P., Scholze, F., Bertelmann, R., Kindling, M., Klump, J., Goebelbecker, H.-J., Gundlach, J., Schirmbacher, P., & Dierolf, U. (2013). Making Research Data Repositories Visible: The re3data.org Registry. *PLOS ONE*, 8(11), e78080. <https://doi.org/10.1371/journal.pone.0078080>
- Prosser, A. M. B., Hamshaw, R. J. T., Meyer, J., Bagnall, R., Blackwood, L., Huysamen, M., Jordan, A., Vasileiou, K., & Walter, Z. (2023). When open data closes the door: A critical examination of the past, present and the potential future for open data guidelines in journals. *British Journal of Social Psychology*, 62(4), 1635-1653. <https://doi.org/10.1111/bjso.12576>
- Shamsudin, Y. M., Zain, W. M., & Sahid, N. Z. (2025). Determinants of Open Data Repository Adoption Among Malaysian Academic Researchers: A conceptual framework. *Environment-Behaviour Proceedings Journal*, 10(SI31), Article SI31.

- <https://doi.org/10.21834/e-bpj.v10iSI31.6932>
- Sharing Study Data: A Guide for Education Researchers / IES.* (s.f.). Recuperado 29 de junio de 2025, de <https://ies.ed.gov/use-work/resource-library/resource/tooltoolkit/sharing-study-data-guide-education-researchers>
- Song, J., Li, C., & Chansanam, W. (2025). Construction of medical scientific data repositories in China: Analysis of survey and recommendations. *Frontiers in Artificial Intelligence*, 8, 1544200. <https://doi.org/10.3389/frai.2025.1544200>
- Strecker, D., Pampel, H., Schabinger, R., & Weisweiler, N. L. (2023). Disappearing repositories: Taking an infrastructure perspective on the long-term availability of research data. *Quantitative Science Studies*, 4(4), 839-856. https://doi.org/10.1162/qss_a_00277
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLoS ONE*, 6(6), e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Thøgersen, J. L., & Borlund, P. (2022). Researcher attitudes toward data sharing in public data repositories: A meta-evaluation of studies on researcher data sharing. *Journal of Documentation*, 78(7), 1-17. <https://doi.org/10.1108/JD-01-2021-0015>
- Ulrich, R., Verburg, M., Priddy, M., Huber, R., L'Hours, H., Neidiger, C., Meijas, G., & Davidson, J. (2024). *M5.5—Initial repository registry support for discovery of repositories, policies, and interfaces.* <https://doi.org/10.5281/ZENODO.10847707>
- UNESCO. (2021). *UNESCO Recommendation on Open Science.* UNESCO. <https://doi.org/10.54677/MNMH8546>
- Vogt, L., Strömert, P., Matentzoglou, N., Karam, N., Konrad, M., Prinz, M., & Baum, R. (2025). Suggestions for extending the FAIR Principles based on a linguistic perspective on semantic interoperability. *Scientific Data*, 12(1), 688. <https://doi.org/10.1038/s41597-025-05011-x>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Witt, M., Cannon, M., Lister, A., Segundo, W., Shearer, K., Yamaji, K., & Research Data Alliance Data Repository Attributes Working Group. (2024). *RDA Common Descriptive Attributes of Research Data Repositories* (Versión 1.0). Research Data Alliance. <https://doi.org/10.15497/RDA00103>
- Zuiderwijk, A., Shinde, R., & Jeng, W. (2020). What drives and inhibits researchers to share and use open research data? A systematic literature review to analyze factors influencing open research data adoption. *PLOS ONE*, 15(9), e0239283. <https://doi.org/10.1371/journal.pone.0239283>

