Revista Latinoamericana de Tecnología Educativa, 24(2)
https://doi.org/10.17398/1695-288X.24.2.123

RELATEC
Revista Latinoamericana de Tecnología Educativa

Received: May 2, 2025
Reviewed: June 20, 2025
Accepted: June 29, 2025

Corresponding author's address:

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

E-mail / ORCID

gonzalo.torterolo@fing.edu.uy

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

rmotz@fing.edu.uy

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

ARTICLE / ARTÍCULO

Open Educational Research Data: Are Repositories Delivering?

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

Gonzalo Torterolo-Silva y Regina Motz

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.

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.

1. Introduction

In today's digital age, the open science movement has profoundly reshaped the paradigms of scientific knowledge production and circulation. Within this transformative framework, public access to research data emerges as a fundamental strategic component, particularly in the field of education. The application of these paradigms not only strengthens the pillars of transparency and reproducibility that underpin the scientific method (OECD, 2020), but also facilitates secondary analysis and contributes to the social impact of knowledge (European Commission, 2016; Wilkinson et al., 2016). The potential benefits are particularly significant in the field of education, where the possibility of comparing results obtained in different sociocultural contexts and pedagogical systems could generate inputs for the evolution of public policies in education.

UNESCO (2021, p. 10) conceptualizes open research data as data that must be «available in a timely manner, in a format that is easy to use, readable, and modifiable by people and machines», as articulated under the FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Wilkinson et al., 2016). The availability of repositories aligned with these principles facilitates the task of researchers, both in terms of depositing and reusing data. However, it is not trivial to decide which repository should be used to deposit, organize, and manage research data. In this regard, the TRUST principles (Lin et al., 2020) represent a valuable reference framework for evaluating the quality of scientific repositories, promoting Transparency (clear usage policies), Accountability (in terms of data collection management), User Focus (availability of tools for exploring, locating, and reusing data), Sustainability (availability of long-term data access services), and Technology (use and implementation of appropriate standards and tools for data management and curation). The comprehensive characterization by Behnke et al. (2020) details how these principles include both organizational requirements (such as clear data management policies and support for standard formats) and technical requirements (from granular metadata to persistent identifier systems).

The organization and grouping of research data collections in repositories responds to different criteria according to which it is possible to broadly categorize repositories as disciplinary (when they contain data from a specific area of knowledge), generalist, and institutional (if they house data from a particular institution). Moreover, in some disciplines, such as economics or epidemiology, most of the data managed is quantitative, while in others, such as the social sciences and education, the data is mostly qualitative. Each of these repository types may require the prioritization of some principles over others, or the use of different standards or technologies for their implementation. In general, as research data from the same discipline share many characteristics, the technological requirements are similar. In this respect, educational research faces specific obstacles, especially when working with qualitative data which, as Antonio et al. (2019) point out, require exceptional levels of contextualization and specialized ethical protection mechanisms, implemented with specific technologies, which many generalist repositories are not prepared to offer. Other as authors, including Strecker, Pampel, Schabinger, and Weisweiler (2023), report on the risks of these repositories in meeting quality standards, further indicating that repositories related to the humanities and social sciences are among the most affected by such problems. Benjelloun, Chen, and Noy (2020) and Gerasimov, Binita, Mehrabian, Acker,

and McGuire (2024) also warn about trends toward centralization of most datasets in generalist repositories. Paradoxically, several recommendations from these generalist repositories encourage researchers to use disciplinary and institutional repositories (Barbosa et al., 2024). In particular, the works of Ávila Barrientos (2024) and Song, Li, and Chansanam (2025) stand out for their comprehensive analysis of guidelines for selecting reliable scientific data repositories. International organizations such as the OECD (2020) and UNESCO (2021), aligned with the FAIR and TRUST principles, have also promoted policies that encourage the use of specialized digital repositories for the storage, preservation, and ethical dissemination of scientific data.

As Borgman, Scharnhorst, and Golshan (2018) argue, these digital infrastructures are no longer mere repositories of information but have become fundamental nodes in contemporary knowledge production networks, facilitating not only access but also interoperability and the progressive enrichment of datasets through successive reuses. Along the same lines, the recent conceptualization by Ávila Barrientos (2024, p. 27) emphasizes their role as dynamic systems that allow for «the storage, organization, retrieval, and access of datasets of diverse thematic and typological nature,» thus constituting essential drivers for open science.

The application of the FAIR (Wilkinson et al., 2016) and TRUST (Lin et al., 2020) principles faces particular challenges in disciplines such as education and the social sciences, where the heterogeneity, sensitivity, and complexity of data make it difficult to standardize and create specialized infrastructures aligned with the requirements of these disciplines. However, the scarcity of disciplinary repositories in these areas, coupled with ethical, legal, and cultural challenges, forces most researchers to rely on institutional or generalist repositories, which do not always meet the specific requirements for optimal management of educational and social data.

As Borgman (2016) warns in his analysis of knowledge ecologies, data only achieve their full scientific value when they are integrated into robust ecosystems that coherently articulate technical, community, and institutional dimensions. It is this articulation that presents particular challenges in the field of education and social sciences, where specialized disciplinary repositories are scarce compared to those pertaining to other areas of knowledge (Krähmer, Schächtele, & Schneck, 2023; Lamb et al., 2024). This disparity is paradoxical when we consider the intrinsic complexity of educational data, which ranges from standardized quantitative measurements to deeply contextualized ethnographic records, often containing sensitive information (e.g., about minors in specific school environments; Gomes et al., 2022).

Faced with these challenges, initiatives such as re3data (Pampel, 2013) have sought to create integrated catalogs that facilitate the location and evaluation of specialized repositories. The informed selection of repositories thus becomes a crucial challenge. In this context, repository catalogs emerge as key tools for searching, filtering, and evaluating options according to technical, ethical, and disciplinary criteria.

However, when it comes to repositories of educational research data, there is very little disciplinary presence. For example, according to information collected in June 2025, the FAIRSharing, DataCite Commons, and re3data catalogs respectively have 1,252, 1,767, and 3,259 repositories registered, only 22, 5, and 47 of which correspond to research data repositories in the field of education.

In this sense, we believe that this work contributes to drawing attention to the needs of education researchers in order to continue improving the service that repositories can offer them, thereby encouraging their development.

Accordingly, in this article, we analyze the need for data publication in education and social sciences, and the characteristics of the repositories most valued by researchers in these disciplines, considering their ethical, technical, and legal concerns. In addition, we explore the role of repository catalogs in the informed selection of repositories, with the aim of strengthening the culture of open data and contributing to the advancement of open science in these disciplines.

2. Methodology

This study was conducted using a sequential exploratory methodological design with the following three interrelated phases: Literature Review, Expert Interview, and Catalog Analysis. Its objective was to identify how repositories respond to the ethical, technical, and legal concerns that researchers have when depositing data from educational research

2.1. Methodological Phases

- Phase 1. Literature Review. Various sources were used: articles in Scopus/WoS (keywords: 'open data' + 'education research' + 'concerns'), as well as UNESCO and OECD reports. Exclusion criteria included: studies on open educational resource repositories and works that focus solely on open access to research publications. For the analysis, assisted coding was applied using Elicit1 to categorize recurring concerns (e.g., privacy, intellectual property, additional effort). The expected outcome was a conceptual map of barriers documented in the literature (2015–2025).

- Phase 2. Expert Interview. The person responsible for the open data repository of the National Agency for Research and Innovation (ANII) of Uruguay was interviewed. The instrument (protocol) was a semi-structured interview lasting 45 minutes. For the analysis, the interview was transcribed and coded in order to confirm or refine the categories identified in the literature.

- Phase 3. Repository Catalog Analysis. A sample of six internationally widely used catalogs was analyzed (OpenDOAR, ROAR, OpenAIRE, FAIRSharing, DataCite Commons). The variables evaluated were: repository activity, creation of persistent identifiers, support for specifying context, compliance with interoperability standards, adherence to licensing policies, and sustainability.

For data triangulation and validation, the three methods: literature review, expert consultation, and catalog analysis, were combined, allowing for the triangulation of information and increasing the validity and reliability of the results.

3. Resultados

In this section, we present the results yielded by the literature review, expert consultation, and analysis of repository catalogs.

3.1. Results of literature review

Although numerous studies have focused on identifying the factors that influence researchers' willingness to share their data openly, only the key works focusing on repositories are highlighted here. A comprehensive review of extant literature reveals that a variety of data collection methods and analytical approaches are adopted in these studies, including surveys, systematic literature reviews, meta-analyses, focus groups, and round tables (Barczak, Hopp, Kaminski, Piller, & Pruschak, 2022; Borycz et al., 2023; Casali, Motz, & Sprock, 2022; Grattarola, Laufer, et al., 2024; Logan, Hart, & Schatschneider, 2021; Mosha & Vincent, 2024; Shamsudin, Zain, & Sahid, 2025; Tenopir et al., 2011; Thoegersen & Borlund, 2022; Zuiderwijk, Shinde, & Jeng, 2020). It is interesting to note that most of these studies were conducted with researchers from different disciplines, with different levels of expertise and from different geographical contexts. Curiously, few of these studies have been studies conducted with education sciences researchers, such as Casali et al. (2022), who analyzed survey responses provided by the LACLO (*Latin American Conference on Learning Technologies*) attendees. On the other hand, Logan et al. (2021) and Neild, Robinson, and Agufa (2022) present guidelines on open data publication specifically aimed at researchers in education.

While the work of Casali et al. (2022) highlights barriers such as the time and effort required to share data, the lack of funding for data standardization, and restrictions related to data security and confidentiality, Logan et al. (2021) and Neild et al. (2022) pay special attention to procedures for maintaining data security and confidentiality, as these are recognized as the most common problem.

Most of the researchers in this field agree that the extra effort involved in curating data for open publication is an important hindrance to the establishment of repositories. While this problem is generally attributed to the development of repositories that are not very user-centric, some authors argue that it is due to a lack of administrative support for researchers in preparing data (Borycz et al., 2023; Casali et al., 2022).

More recently, Grattarola, Shmagun, et al. (2024) and Mabile et al. (2025) pointed out that the main disincentive to sharing data openly is the lack of institutional incentives and recognition for data publication. With regard to qualitative research data, all authors highlight the problem of having elements that allow the context in which the data were obtained to be specified and the concern for having specific curation to ensure researchers' protection from the legal and ethical problems arising from using sensitive data.

Qualitative researchers in the social sciences mention issues related to the privacy of their data and the ethical commitment involved in its publication. In the social sciences, data are usually provided by or relate to study participants, and even if they are anonymized, many researchers fear that participating individuals may be re-identified, especially when data was collected in small communities or relates to sensitive topics (Krähmer et al., 2023). Despite the existence of mechanisms to mitigate this issue, such as informed consent, the responsibility of repositories for data custody is not always easy for data producers to understand (Prosser et al., 2022).

In their work, Lamb et al. (2024) confirm some of the challenges inherent in qualitative research. According to their findings, qualitative data (such as interviews or field diaries) require in-depth contextual knowledge to be interpreted properly when other researchers decide to reuse the dataset.

3.2. Results of expert consultation

The conclusions drawn from the expert interview are based on an internal study conducted by the National Agency for Research and Innovation (ANII). Although unpublished, this study provided valuable information on researchers' perspectives in the context of Uruguay. The ANII study revealed that researchers tend to perceive themselves primarily as consumers of data rather than producers. This perception influences their expectations and needs in relation to data repositories. Discussing the findings yielded by this study, the expert interviewed highlighted three key characteristics that researchers consider essential when publishing their data:

a) Assignment of persistent identifiers: The assignment of unique and permanent identifiers, preferably Digital Object Identifiers (DOIs), is seen as a fundamental requirement, especially since many scientific journals require them for the evaluation and citation of articles.

b) Sustainability and durability: The repository's ability to guarantee long-term data preservation is considered crucial to ensuring its availability and future reuse.

c) Interoperability: The repository's ability to facilitate data integration and exchange between different platforms and systems is considered essential for maximizing the impact and reuse of data.

The semi-structured nature of the interview facilitated in-depth exploration of relevant topics without restricting the spontaneity of the dialogue, which helped to validate and expand the results of the bibliographic survey with empirical evidence and direct experience.

3.3. Results of the repository catalog analysis

As a part of their studies, Witt et al. (2024) and Song et al. (2025) established the main criteria for evaluating research data repositories. However, these authors point out that researchers face obstacles in verifying compliance with these criteria, given the limited accessibility of this information a priori. This limitation makes it difficult to select a suitable repository for the publication and reuse of data. Guided by this observation, in this paper, we analyze the role of repository catalogs in the efficient dissemination of such information, taking as a reference the criteria defined by these authors and those developed by the Data Repository Attributes Working Group 1. The characteristics we seek to identify are: repository activity, creation of persistent identifiers, support for specifying context, compliance with interoperability standards, and adherence to licensing and sustainability policies.

Repository activity

The activity factors of the repository and disciplinary specificity strongly determine the visibility and reuse of data. Recent repository activity is a key indicator of its reliability. Indicators of activity include the frequency of updates (e.g., new datasets deposited

monthly), date of the last dataset deposited (researchers should avoid repositories with no activity in the last 12 months), and public statistics (e.g., number of downloads or citations in DataCite).

On the other hand, a repository specializing in a subject area offers the significant advantage of using specific metadata schemas (e.g., DDI[1] for social sciences), which improves interoperability and data discovery. In addition, thematic repositories attract specialized users, which increases the impact and citation of stored datasets.

Creation of persistent identifiers

The DOI plays a fundamental role in research data repositories, as it guarantees the location, citation, and permanent preservation of datasets. It provides a unique and stable link, preventing data loss due to changes in URLs or outdated repositories. This ensures that datasets remain accessible in the long term, complying with FAIR principles. By integrating with bibliographic management systems (such as Crossref or DataCite), the DOI allows datasets to be linked to scientific articles, improving their visibility and facilitating the tracking of their impact.

The DOI is assigned by authorized registration agencies, with DataCite being the most widely used for research data or by repositories/institutions affiliated with DataCite. In some cases, the same dataset may be associated with multiple DOIs. This can happen if a researcher uploads the same dataset to different repositories and each one assigns it a different DOI, or in the case of updates or versions, as some repositories generate a new DOI for corrected or expanded versions of the same dataset. This practice is undesirable as citations to the dataset are scattered across multiple DOIs, underestimating its real impact and making it difficult to identify which is the most up-to-date version. To avoid these problems, it is recommended to prioritize the use of reference repositories in each discipline (e.g., ICPSR in social sciences) to avoid duplication. If this is not possible, using repositories that support versions under the same DOI and linking alternative DOIs using metadata (isVersionOf or isIdenticalTo relationships in schemas such as Schema.org) is advised.

Support for documenting context

The context of a dataset includes details about its origin, the methodology used to generate it, its potential uses, and its limitations. Repositories collect contextual information through descriptive metadata, technical documentation (such as laboratory protocols, questionnaires, or processing scripts), additional documents (such as readme files, guides, and reports), and linked publications (DOIs of related articles in which the dataset was used or cited). The descriptive metadata must include at least the following information:

a) Origin: Who created it (institution, research project, authorship) and temporal/geographical coverage (when and where the dataset was generated).

b) Purpose: Objective of the collection.

c) Collection methods: Instruments, protocols, or software used.

d) Licenses and restrictions: Conditions of use (CC-BY, GDPR, restricted access).

[1] https://ddialliance.org/

Catalogues such as OpenAIRE and DataCite Commons promote the use of international standards (such as the OAI-PMH protocol and FAIR metadata), which ensures that data is easily locatable, accessible, and reusable by different platforms and systems, fostering collaboration and information exchange between scientific communities. This best practice of using controlled vocabularies and disciplinary ontologies (e.g., OECD Education Taxonomy) to facilitate interoperability is promoted by FAIRSharing, which offers a curated directory of metadata standards among its features. When a repository is indexed in FAIRSharing, it evaluates and recommends specific metadata standards according to its discipline.

Compliance with policies

Another aspect to consider in a repository, which depends on political or legal factors, is its usage and preservation or sustainability policies. The problem of repositories becoming extinct is highlighted by Strecker et al. (2023). Some catalogs, such as FAIRSharing, facilitate the evaluation of these factors, allowing the relationship between the specific policies of each repository and more general policies to be identified.

If we take the example of the NIH[2] as a funder, it could impose compliance with its terms of data use[3]. FFAIRSharing is capable of displaying repositories that are compatible with this policy, as depicted in Figure 1. Aided by this functionality, a researcher with an NIH-funded project can limit the selection of repositories to those found in the query that produces the graph in Figure 1.

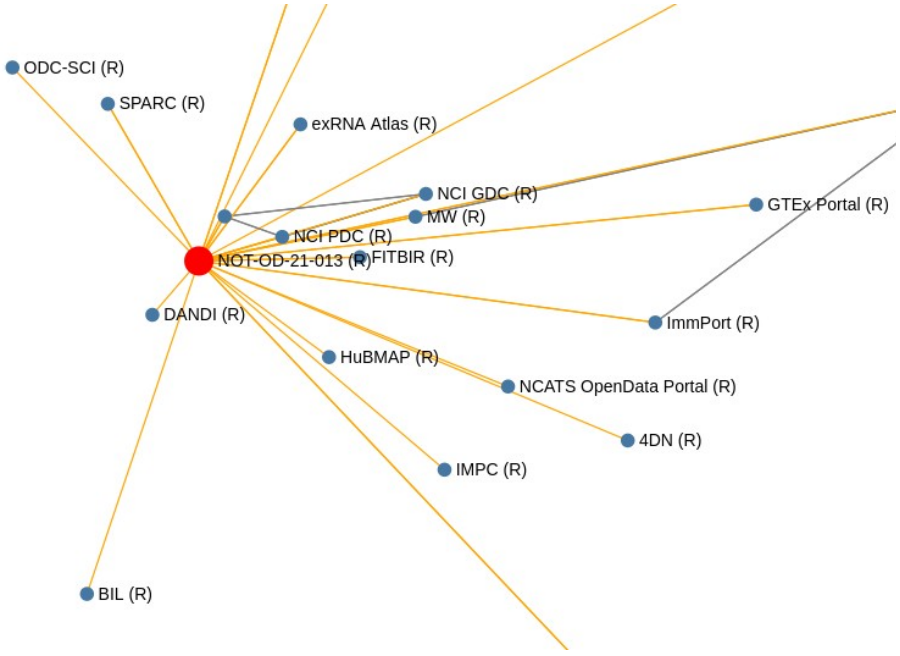

Figure 1. Graph of repositories that use the NIH policy – FAIRSharing (Retrieved on 2025/06/26).

[2] https://www.nih.gov/
[3] nih_policy_fairsharing] 2023 Final NIH Policy for Data Management and Sharing - https://fairsharing.org/FAIRsharing.7861ef

Another mechanism that helps identify repositories' adherence to certain usage and sustainability policies is certification. Repository certifications are external evaluation mechanisms that guarantee that a repository complies with quality, preservation, and best practice standards in data management. Certifications are especially useful for determining whether repositories adhere to preservation or sustainability policies. For instance, a certification indicates that a repository has demonstrated some of the following functionalities: it assigns DOIs, uses disciplinary metadata schemas, has a sustainable technical infrastructure (backups, format migration), offers version management for change control in datasets, has technical support to respond to incidents, and is backed by a stable institution (university, public body). Different certification seals signify different functionalities. For example, a repository with CoreTrustSeal[4] seal confirms that a repository has undergone a thorough verification of legal compliance criteria (e.g., GDPR, intellectual property) but not of DOI assignment.

The catalogs compile data from certifying bodies and standardize the display of certifications to facilitate comparison. For example, re3data shows the year of certification and its validity. FAIRSharing does not certify, but rather evaluates and recommends repositories that adhere to the use of metadata standards and FAIR principles. It also includes a filter by certification (e.g., CoreTrustSeal, ISO 16363[5]) and provides links to official certification reports. A researcher looking for a repository for education data can use FAIRSharing and filter by CoreTrustSeal, as well as check in re3data to see if the repository has a valid certification.

Repository compatibility with services

Knowing the repository's capacity to provide or integrate with certain specialized services in a particular research area, as well as the conditions under which it provides these services (price or level of availability), may be relevant to researchers. Some of these services and functionalities include the ability to perform or receive formal reviews, the possibility of establishing embargo periods, specialized technical assistance, or the availability of disciplinary curation services.

In particular, in the case of qualitative research data repositories, the presence or integration of services such as anonymization or access control mechanisms is valuable to researchers. Catalogues such as OpenAIRE provide information on the integration of different repositories with the EOSC service ecosystem. For example, using the «Compatibility Level» filter, researchers can view the repository's level of compatibility with services such as Amnesia[6] nonymization service, OpenAIRE AAI[7], API for access service, and «Argos»[8] Data Management Plan service, among others. Table 1 provides a summary of the main features of the repository catalogs reviewed.

[4] https://www.coretrustseal.org
[5] https://www.iso.org/standard/87472.html
[6] https://www.openaire.eu/amnesia-guide
[7] https://catalogue.openaire.eu/service/openaire.aai/overview
[8] https://catalogue.openaire.eu/service/openaire.argos/overview

Table 1. Main features of six repository catalogs in focus of this work.

Catalog	Features
OpenDOAR	Provides basic information about the repository.
ROAR	Similar to OpenDOAR, but allows users to find out the discipline associated with the repository. Specializing in academic publication repositories, it also contains research data repositories on an occasional basis.
Re3data	Specializes in describing research data repositories. Provides the largest number of repository records. Its publication on certification validity is very useful.
FAIRsharing	It allows users to navigate between the relationships between standards and policies and different data repositories. Standards include FAIR principles, TRUST, and various certifications. It includes various repositories (data, documents, multimedia, software, etc.).
DataCite Commons	It integrates information on various types of research repositories and stands out for providing concise information on their content, although it does not allow users to explore certain characteristics of the repository in detail, nor does it offer advanced filtering criteria like re3data.
OpenAIRE	Similar to DataCite, it allows users to view, but not navigate between, repository relationships, making it less useful than FAIRSharing. It stands out for the quality with which it handles data provenance, allowing users to see which repositories share content and which catalog the repository information comes from.

4. Conclusion

The following section presents the findings obtained from the results of the applied methodology. The findings are summarized in Table 2. The first column shows the characteristics of repositories that meet the needs identified by researchers listed in the second column, and the third column lists the catalogs that, to a greater or lesser extent, provide information that allows researchers to identify which repositories meet those characteristics.

Table 2. Features of repositories most valued by researchers.

Feature	Researchers' needs	Catalog providing information
Activity, Reputation	Recognition	Re3data, FAIRSharing, OpenAIRE
Technical support	Less effort	FAIRSharing, DataCite Commons
Metadata availability	Context information	FAIRSharing, DataCite Commons, OpenAIRE
Compliance with FAIR principles	Data access and reuse, interoperability	Re3data, FAIRSharing, OpenAIRE
Sustainability, Policy Compliance	Data availability and integrity	Re3data, FAIRSharing

The features presented in the first column of the Table 2 represent the attributes of repositories that are identified as most valued by the study authors, as shown below:

a) Activity and Reputation: The repository must comply with disciplinary community standards, have recognized activity and enjoy a good reputation within the scientific community, and be supported by professional organizations or academic institutions.

b) Technical support: The repository must provide comprehensive documentation and good technical support to facilitate its use.

c) Metadata availability: The repository must contain descriptive metadata that provides contextual information about the data, such as the author and purpose of the collection, and the methodology used for its compilation and maintenance. (

d) Compliance with FAIR principles: The repository must provide clear metadata and support interoperability between different data formats. It must comply with a clear governance structure and transparent policies that enable secure and ethical reuse.

e) Sustainability and policy compliance: The repository must have a long-term financing model and a plan to ensure its continued operation, and it must include continuity plans to prevent loss of information. The repository must have a solid infrastructure in terms of storage capacity, data security, and backup protocols.

4.1. Discussion

It is important to highlight researchers' openness to publishing data and their appreciation of the benefits this brings to both those who publish and those who access the data. However, current data repositories do not seem to offer a comprehensive solution that meets all the needs of educational researchers.

Despite this limitation, the work presented here provides a comparative evaluation of research data repository catalogs. Based on the obtained findings, FAIRSharing stands out as the tool that offers the best levels of curation and data quality. Its sustainability model and the transparency of its criteria and processes are also noteworthy. FAIRSharing provides an organic view of the publication process and open research infrastructure. However, its exclusive use is not recommended. Because the FAIRSharing database lacks some records, researchers should perform a parallel query in DataCite Commons, and should use re3data to obtain a better view of the validity of certifications.

Given the limitations that hinder visibility, several authors of extant studies in this domain concur that it is necessary to improve the discoverability of available repositories and services, as recently pointed out by Huber, Gonzalez Beltran, Neidiger, Ulrich, and L'Hours et al. (2024). These authors also recognize the need to improve the capacity of repositories and infrastructure to capture diverse disciplinary information. Other researchers, including Ulrich et al. (2024), propose mechanisms for discovering the services and capabilities of each repository, even though this approach still requires standardization efforts on the part of repository managers to be useful. Implementing

such proposals in practice would greatly enhance the quality of the data collected by repository catalogs, making them much more reliable and descriptive.

Some researchers, for instance Vogt et al. (2025), choose to neglect the role of repositories and focus instead on content, proposing the creation of a layer of semantic interoperability that unifies resources from a conciliatory perspective for each discipline, facilitating access by researchers not only to research data but to resources in general. According to these authors, this approach blurs the line that differentiates research articles from research data or even resources external to the research area (e.g., OERs, government data, software, etc.), concepts that are difficult to differentiate for the majority of researchers and that often may not make sense to differentiate from the researcher's point of view.

Other proposed strategies include the use of data management plans (DMPs) to make the dialogue between the different actors during the research process transparent (Bicarregui et al., 2012). Miksa, Mietchen, and Simms (2019) go a step further to suggest capturing the dialogue in a structured and automatically actionable document (maPGD). Although the use of DMPs is not new, their adoption still needs to be increased before they can be fully exploited.

Such initiatives will help bridge the disciplinary gap and promote dialogue between data management and disciplinary knowledge.

5. References

Antonio, M. G., Schick-Makaroff, K., Doiron, J. M., Sheilds, 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 informacion*, *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., Martinez, 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., Giaretta, 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

Thoegersen, 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., Matentzoglu, 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, Ij. 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