

Criterios para selección de repositorios para datos abiertos de investigación

Gonzalo Torterolo
Regina Motz, Edelweis Rohrer



Resumen

- Definir el objeto de estudio
- Identificación de necesidades
- Rol del repositorio (funcionalidades)
- Catálogo de repositorios
- Comparativa de catálogos de repositorios



“Datos” de investigación

El objeto de estudio son los repositorios, pero es un concepto un tanto esquivo o amplio, incluso su contenido (los datos).

- Ejemplos:
 - Mapas, bases de datos, modelos semánticos, entrevistas, fotografías, etc. (contenidos)
 - Imágenes, vídeos, tablas, textos, etc. (formatos)
- Son datos de investigación?
 - Objetos físicos.
 - Cuadernos de laboratorio.
 - Parámetros del proceso metodológico aplicado.
 - Programas de software.
- Metadatos
 - Es un término bastante polémico, puesto que depende del uso que se esté dando al dato.
 - En este contexto, se refiere a datos de gestión.

Datos “Cualitativos”

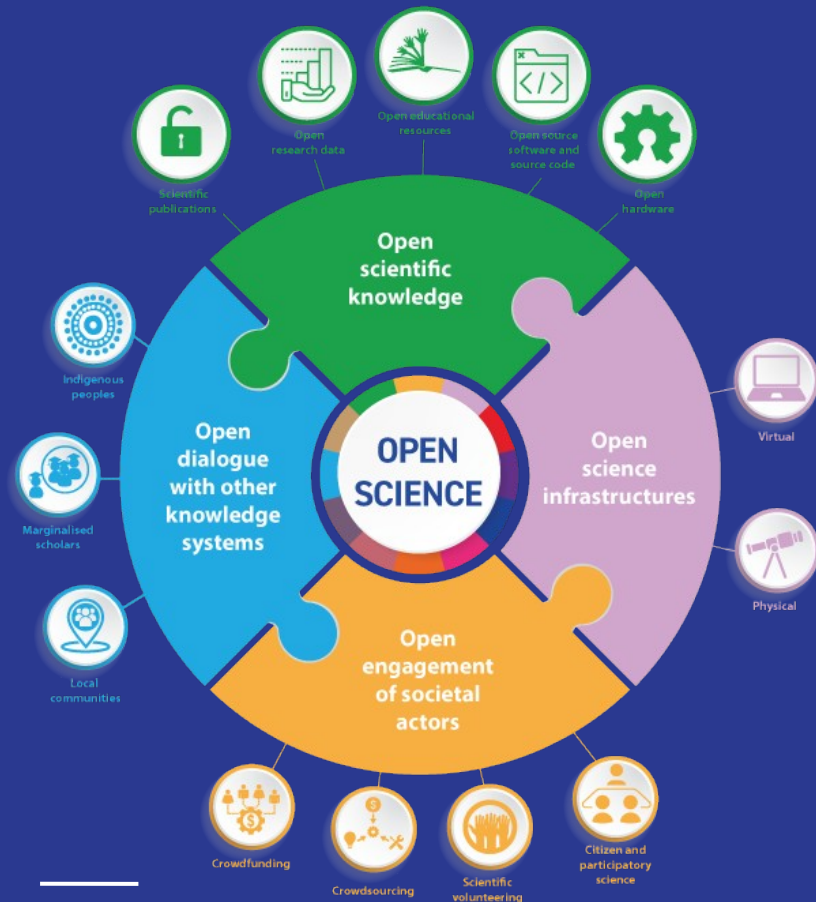
- Los datos no son cualitativos, sino más bien el método con el que se produjeron.
- Tipos de datos que suelen estar asociados a estudios cualitativos:
 - Opiniones de los clientes sobre un producto o servicio.
 - Motivaciones de los usuarios para realizar una determinada acción.
 - Experiencias de los pacientes durante un tratamiento médico.
 - Comportamientos de los estudiantes en el aula.
 - Resultados de entrevistas en profundidad sobre un tema específico
- Particularidades:
 - Contextuales -> Parámetros. Limitar el contexto o permitir conceptos emergentes?
 - Heterogéneos -> Multiplicidad de formatos
 - Restricciones éticas
- En la medida de lo posible, nos vamos a enfocar en el área de las ciencias de la educación.

Datos abiertos en ciencia abierta

“La ciencia abierta redefine la producción y circulación del conocimiento, facilitando el acceso al saber. Este enfoque permite que los datos sean accesibles y reutilizables, impulsando la colaboración entre investigadores y la sociedad en general.”

“El acceso público a los datos de investigación es crucial para la transparencia y la reproducibilidad. Esto permite a los investigadores compartir sus hallazgos, fomentando la innovación.”

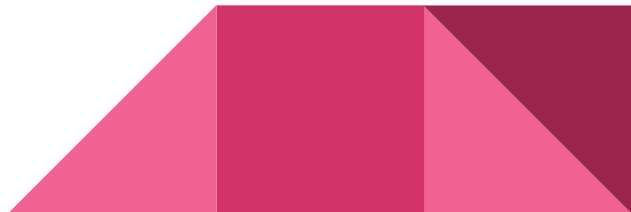
Todo ésto no se hace solo...



Utilidad de compartir datos

Son un elemento fundamental y deben ser publicados, pero requieren una gestión y curado particular, diferente a otros elementos.

- Preservación y memoria institucional.
- Transparencia y Reproducibilidad
 - Toma de decisiones.
 - Proceso de validación por pares.
- Reutilización y aumento de la eficiencia.
- Impacto social y público.
 - Difusión a la sociedad.
 - Aplicación fuera del ámbito científico.
- Reconocimiento directo (citas) o indirecto.



Barreras para la publicación de datos

Costo-Beneficio de las prácticas de publicación de datos.

Dilema de marco de aplicación.

Dilema de valoración de los datos.

Poco apoyo o incentivo institucional.

Falta de formación en gestión de datos.

Preocupaciones sobre aspectos técnicos y legales.

Aspectos tecnológicos

Ausencia de estándares técnicos comunes.

Escasa interoperabilidad y pobre experiencia de usuario.

Repositorios especializados que contemplen la sensibilidad de los datos cualitativos.





Criterios de calidad de datos - FAIR

● Encontrable

- Los metadatos siempre están disponibles y permiten saber que la información existe, y que acciones tomar para acceder a ella.
- La información que no se encuentra o no está disponible en el lugar adecuado, es como que no existiera, desperdiciando gran parte del esfuerzo de la investigación.

● Accesible

- Tan abierto como sea posible tan cerrado como sea necesario.

● Interoperable

- Permite el acceso tanto por humanos como por máquinas.
- Permitir moverse libremente entre diferentes herramientas.

● Reutilizable

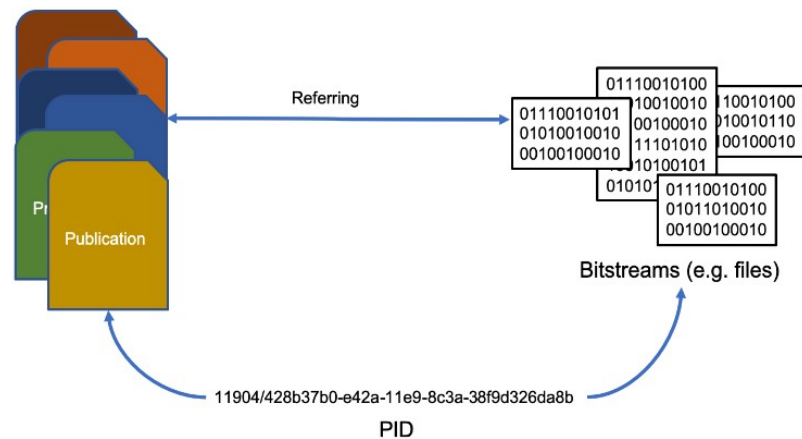
- Atribución y citado.
- Modo de reutilización claro.
- Reutilización responsable.





Criterios de calidad de datos - FAIR

- **Preservación**
 - Sustentabilidad.
 - Archivado duradero y bajo condiciones establecidas.
 - Gestión activa para evaluar la vigencia y acceso de los materiales archivados
- **Transparencia**
 - Procedencia.
 - Reproducibilidad.
- **Reutilización**
 - Atribución y citado.
 - Reutilización responsable.
- **Difusión**
 - En las comunidades científicas adecuadas.
 - También fuera del ámbito académico.
- **Reconocimiento**
 - El autor es reconocido mediante un proceso claro.
 - A los consumidores les es más fácil contactar con los productores de los datos.



Repositorios

- La información provista por los investigadores al publicar sus datos es el insumo principal para la gestión de los datos.
 - Metadatos adecuados potencian la reutilización responsable.
- Las instituciones, las políticas, las agencias y el entorno todo deben colaborar o evitar complicaciones innecesarias.
 - Las exigencias son cada vez mayores.

NO es responsabilidad de una sola persona!



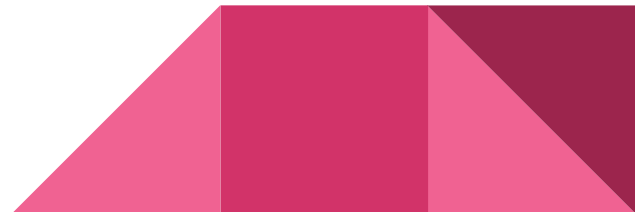
Repositorios

- La tecnología tienen que facilitar este proceso.
- Existe una brecha metodológica entre los diferentes actores de un repositorio que ameritan un espacio de diálogo.
- Los repositorios son un articulador tecnológico entre los diferentes actores del entorno.
- Es difícil delimitar las responsabilidades del repositorio.
 - Otros componentes son sistemas CRIS, catálogos semánticos.
 - Cuando funciona bien, es invisibilizado.



Donde depositar? Repositorios generalistas

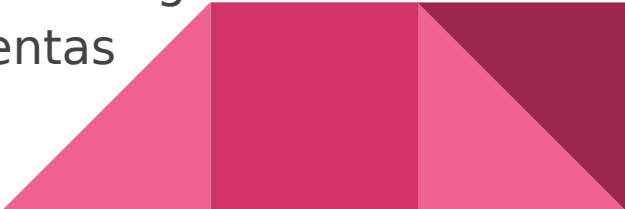
- Web Of Science, [Academia.edu](https://www.academia.edu), ResearchGate
 - Incluyen funciones de repositorio, pero incentivan prácticas no alineados con la ciencia abierta.
 - No está claro el licenciamiento de los datos publicados allí.
- Existen varias plataformas populares
 - Generalist Repository Ecosystem Initiative: GREI.
 - Zenodo, MendeleyData, Dataverse, OSF, Figshare, Dryad, DataHub.
 - ScienceDataBank, 4tuResearchData
- Confiables y generalmente sustentables, pero garantías best-effort.
- Muchas veces satisfacen los requisitos de una agencia de financiación.



Repositorios generalistas: Limitaciones

- Falta de especialización temática
 - No están diseñados para abordar la complejidad ni la sensibilidad de los datos educativos o sociales.
 - Carencias en curaduría disciplinar de datos.
- No ofrecen mecanismos específicos de protección ética (licenciamientos específicos, embargos).
- Falta de herramientas de contextualización profunda.

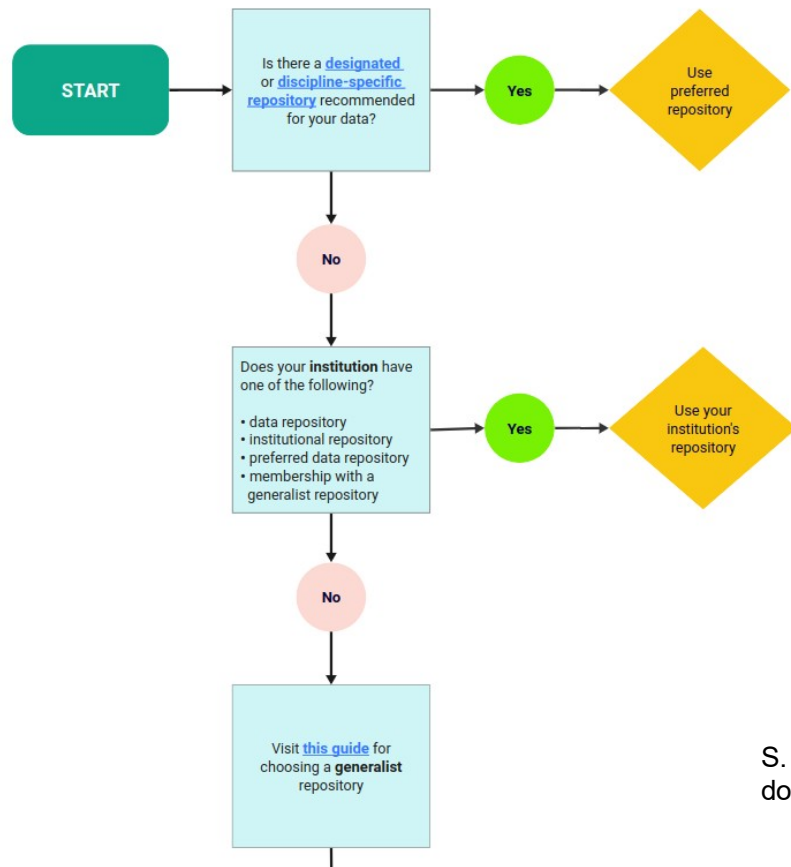
Frente a estos estos escenarios, los mismos repositorios generalistas suelen recomendar el uso de otro tipo de herramientas



Repositorios generalistas: Limitaciones



Generalist Repository Ecosystem Initiative



Each repository shares the following features:

- version support
- use metrics (**Make Data Count**)
- embargo option
- peer review access
- metadata schemas
- JSON for data export
- PIDs for data creators/authors, identifiers & affiliations, funders
- grant ID(s) & grant ID affiliation(s)
- linking to related software and research works/objects
- **FAIR compliant**

Repository	Identifier	Links to submission guidance and requirements, FAQs & user guides
Dataverse		https://guides.dataverse.org/en/5.9/user/
Dryad		https://datadryad.org/stash/requirements
Figshare		https://help.figshare.com/section/user-guides
Mendeley Data		https://data.mendeley.com/faq#mendeley-data
Open Science Framework (OSF)		https://help.osf.io/article/342-getting-started-on-the-osf
Vivli		https://vivli.org/resources/resources/
Zenodo		https://help.zenodo.org/faq/

S. Barbosa, “Generalist Repository Selection Flowchart”. Zenodo, May 17, 2024.
doi: 10.5281/zenodo.11105430.

Repositorios disciplinares e institucionales

● Institucionales

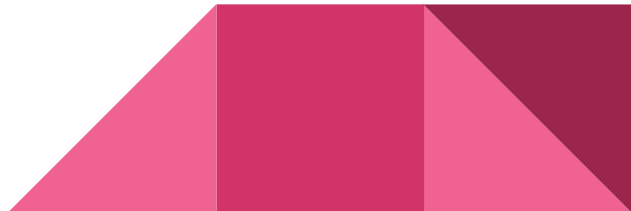
- Garantías de calidad de servicio.
 - Respaldo de datos.
- Apoyo técnico e institucional directo.
 - Económico. Curaduría especializada.
- Cumplimiento con políticas locales.
 - Adaptación a licenciamientos locales y aprobación.

● Disciplinares

- Metadatos y formatos estandarizados según la comunidad.
- Mayor visibilidad entre pares.
- Revisión curatorial especializada.

● Un poco de ambas

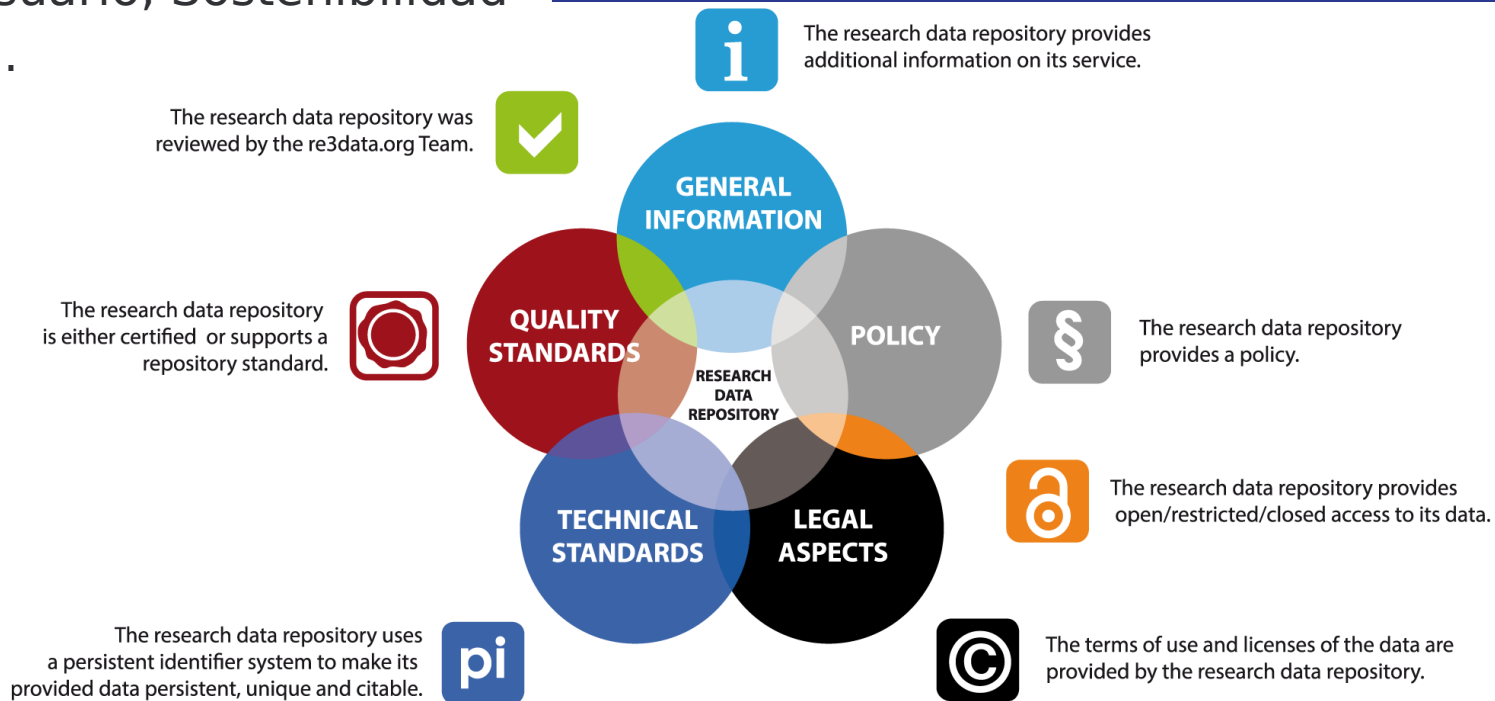
- Multidisciplinaridad.



Evaluación de repositorios

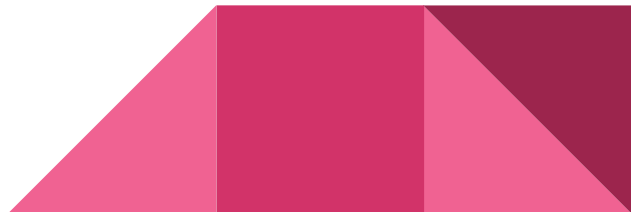
- Principios TRUST

- Transparencia, Responsabilidad, foco en el Usuario, Sostenibilidad y Tecnología.



Evaluación de repositorios - Principios TRUST

- TRUST evalúa la evolución, confiabilidad y preservación de los datos, y los procesos (políticos, económicos, técnicos) que la garantizan.
- Es complementario a FAIR y mayormente se enfocan solo en las características de los repositorios y no en su contenido.



Evaluación de repositorios - Principios TRUST

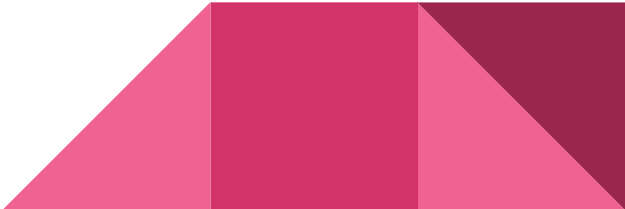
● Sustentabilidad:

- Políticas institucionales y programas de financiación deben incentivar la publicación de datos.
- Modelos de financiación claros.

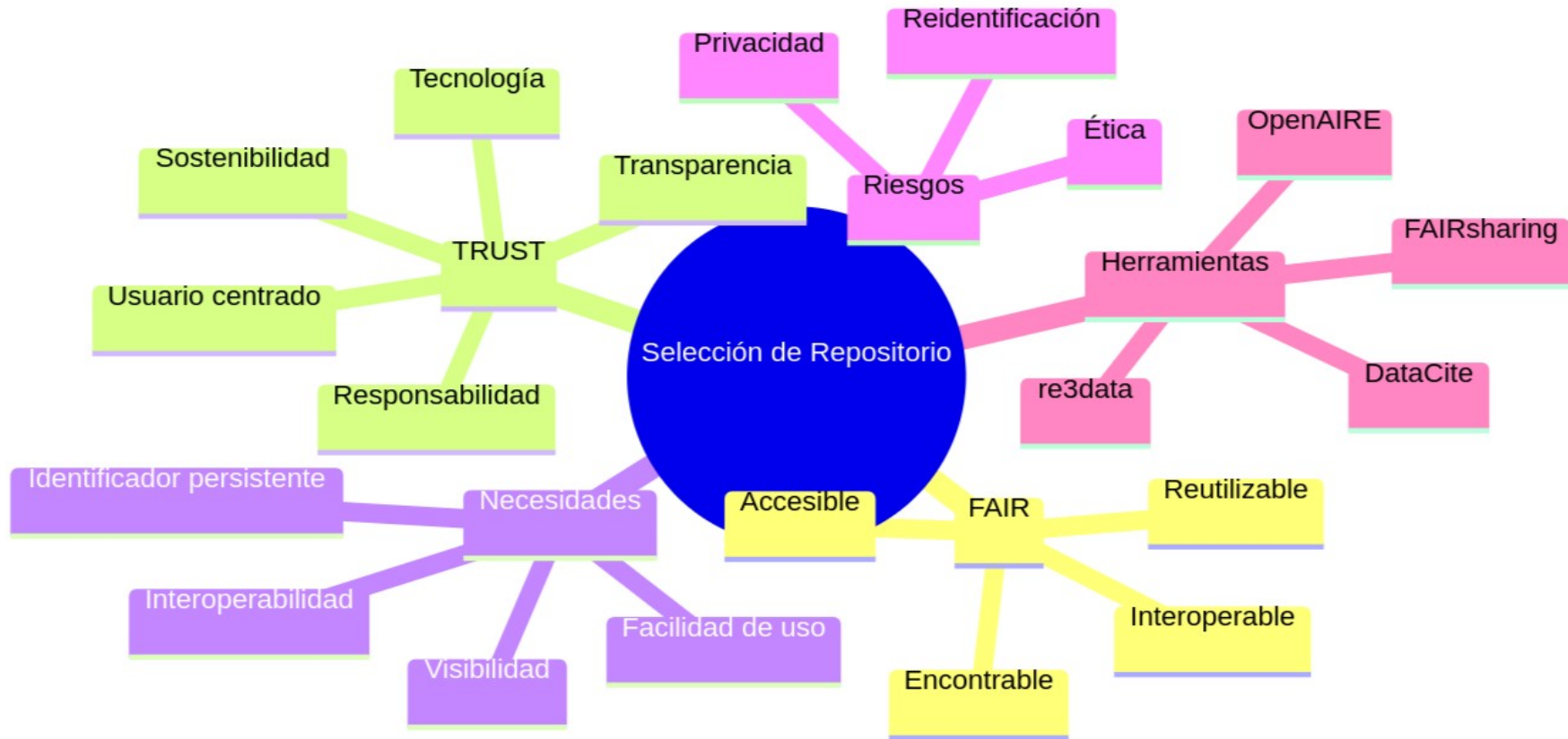
● Interoperabilidad:

- Reutilizados entre diferentes plataformas.
- Un repositorio que garantice esta capacidad maximiza el impacto de las investigaciones al facilitar el intercambio de datos entre investigadores y disciplinas.

● Foco en el usuario:

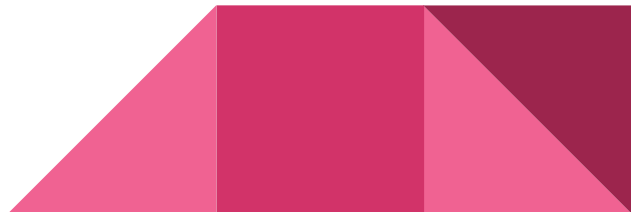
- Los investigadores en educación requieren repositorios que aseguren la disponibilidad de metadatos claros y bien estructurados.
 - Apoyo técnico.
 - Documentación.
 - Facilidad de uso.
- 

Elementos a considerar durante la evaluación



Ventajas de utilizar catálogos de repositorios

- Los catálogos de repositorios es una herramientas clave para evaluar y comparar diversas opciones de publicación de datos.
- Permiten filtrar e identificar los repositorios que cumplen con criterios técnicos, éticos y disciplinares relevantes.
- Evitar tener que navegar en cada repositorio para descubrir sus características.
- Nuclea opiniones y experiencias de otros usuarios sobre el repositorio.



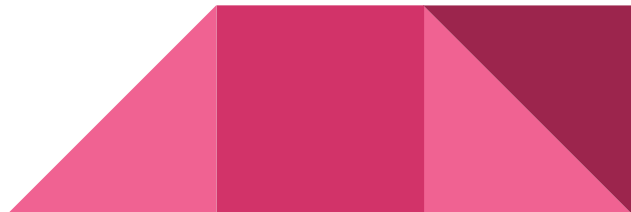
Catálogo de repositorios

- OpenDOAR
- ROAR
- Re3Data
- OpenAIRE
- FAIRsharing
- DataCite Repository Finder

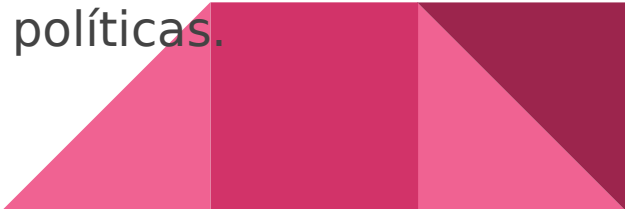


OpenDOAR y ROAR

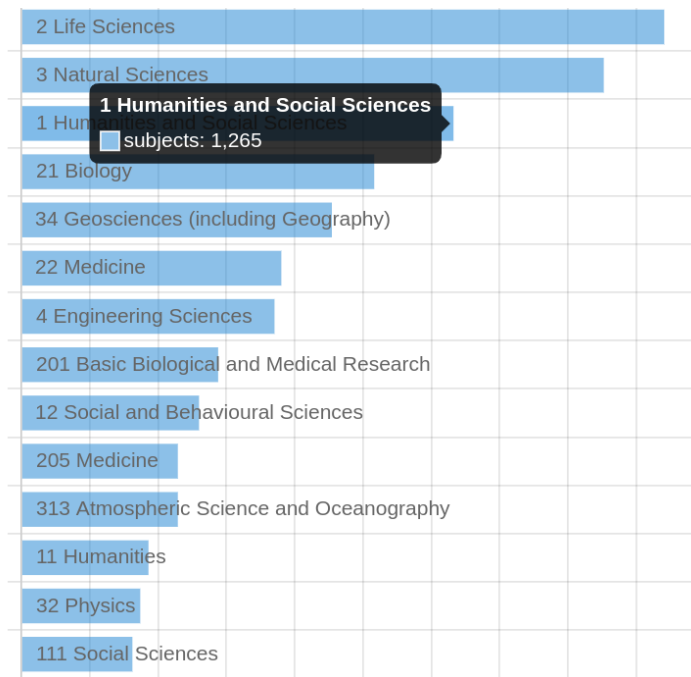
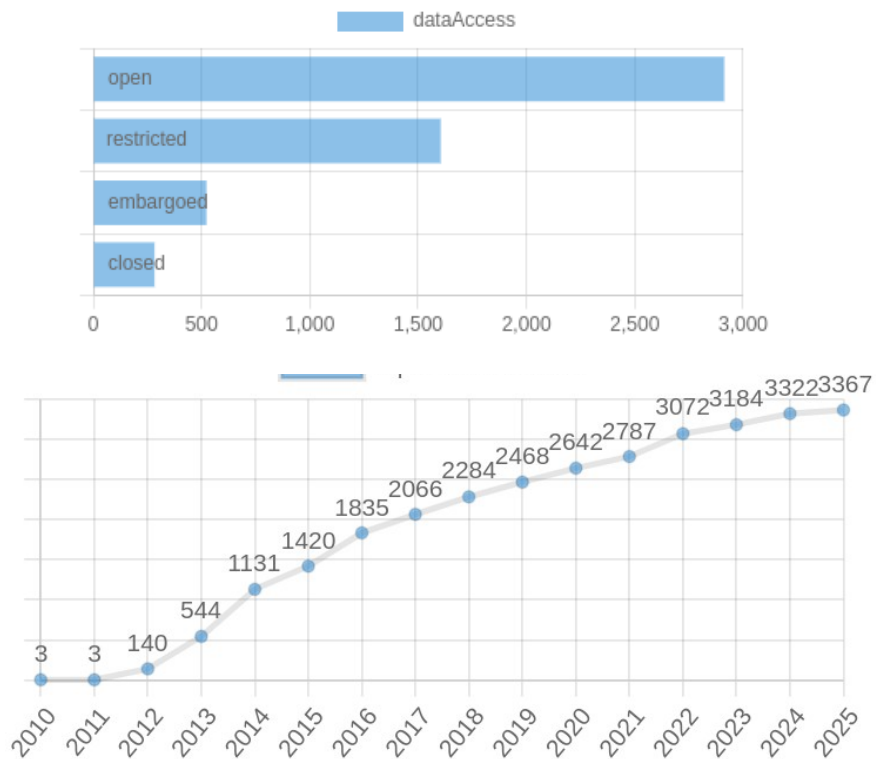
- Largamente utilizados, sobre todo para el registro de repositorios institucionales de acceso abierto (artículos, tesis, etc).
- No proporcionan datos con suficiente especificidad para un análisis automatizado en profundidad.
- Pobre experiencia para usuarios no gestores (eg: investigadores).



Re3Data

- Se identificó como la fuente de datos primarios más completa y actualizada sobre repositorios de datos de investigación.
 - Presenta algunos problemas sobre la trazabilidad y actualidad de sus registros.
 - Varios registros presentan errores de normalización subsanables automáticamente.
 - La experiencia de usuario está acotada brindar información sobre los repositorios de datos y no permite navegar fácilmente entre otros elementos como autores, instituciones o políticas.
- 

Re3Data - Estadísticas



Repository details



Redata

General Institutions Terms Standards

Institutions (2)

Institution	Agencia Nacional de Investigación e Innovación
Additional name(s)	ANII
Institution	ROR:03egaj678
Country	Uruguay
Type(s) of responsibility	general
Type of institution	non-profit
Institution	Dataverse Project
Institution	RRID:SCR_001997
Country	International
Type(s) of responsibility	technical
Type of institution	non-profit

[Back to search](#)

[Submit a change request](#)

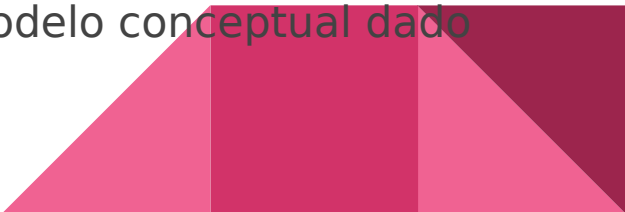
[Get a badge](#)



Cite this re3data.org record:

re3data.org: Redata; editing status 2025-03-07; re3data.org - Registry of Research Data Repositories.
<http://doi.org/10.17616/R31NJNL8> last accessed: 2025-05-14

FAIRsharing y OpenAIRE

- Proporcionan una visión orgánica del proceso de publicación y de la infraestructura de investigación abierta, permitiendo navegar entre políticas, instituciones, estándares, etc.
 - Resulta más sencillo de utilizar por usuarios finales.
 - En este sentido, FAIRSharing se destaca como la herramienta que ofrece los mejores niveles de curaduría y calidad de datos.
 - FAIRSharing destaca por su modelo de sustentabilidad (colaborativo, comunitario, alineado con RDA), la transparencia de sus criterios y procesos.
 - OpenAIRE provee una buena integración con el resto de la nube de EOSC (Anonimización, PGD, PKI, etc) así como asume el modelo conceptual dado por esta.
- 

Search through current results.

MATCH ALL TERMS MATCH ANY TERM

MAINTAINED NOT MAINTAINED

RECOMMENDED NOT RECOMMENDED

READY DEPRECATED UNCERTAIN IN DEV.

Registry

Record Type

Subjects



INTERMAGNET

International Real-time Magnetic Observatory Network

The INTERMAGNET vision is of a global, real-time, permanent geomagnetic observatory network, which is recognized as a key earth observation system and which provides data that serves scientific...

Linked Collections: 0
Linked Databases: 0
Linked Policies: 0



GeoReM

Geological and Environmental Reference Materials

GeoReM is a Max Planck Institute database for reference materials of geological and environmental interest, such as rock powders, synthetic and natural glasses as well as mineral, isotopic, biologic...

Linked Collections: 0
Linked Databases: 1
Linked Policies: 0



zenodo

Zenodo (Zenodo)

Repository

Database

Zenodo is a generalist research data repository built and developed by OpenAIRE and CERN. It was developed to aid Open Science and is built on open source code. Zenodo helps researchers receive credit by making the research results citable and through OpenAIRE integrates them into existing reporting lines to funding agencies like the European Commission. Citation information is also passed to DataCite and onto the scholarly aggregators. Content is available publicly under any one of 400 open licences (from opendefinition.org and spdx.org). Restricted and Closed content is also supported. Free for researchers below 50 GB/dataset. Content is both online on disk and offline on tape as part of a long-term preservation policy. Zenodo supports managed access (with an access request workflow) as well as embargoing generally and during peer review. The base infrastructure of Zenodo is provided by CERN, a non-profit IGO. Projects are funded through grants.

<https://www.zenodo.org>

2013

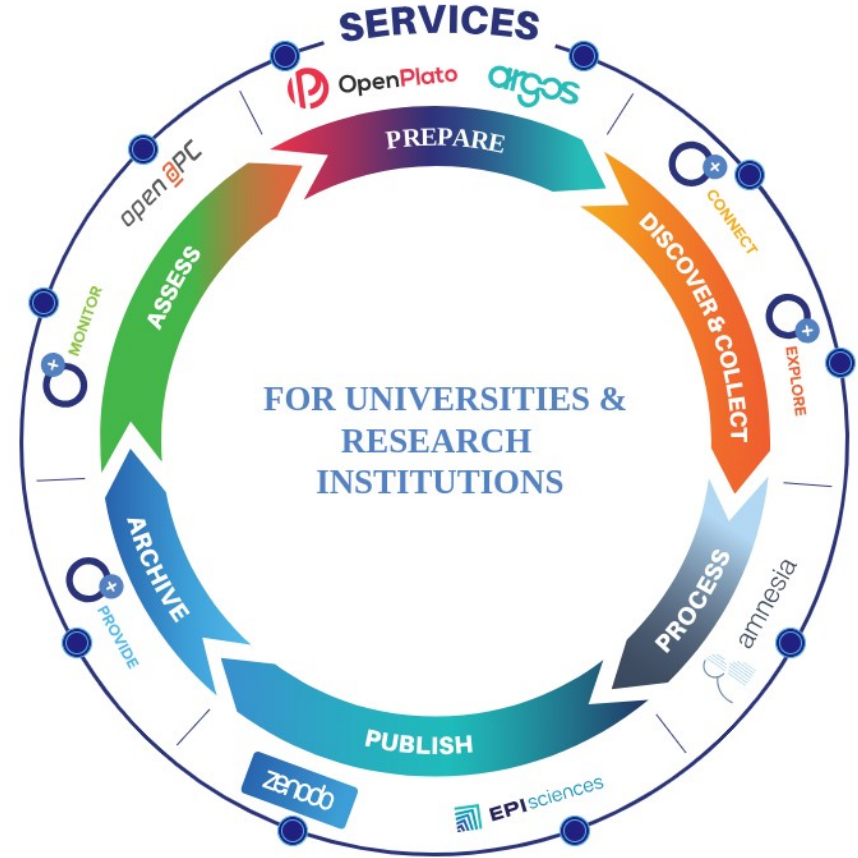
[zenodo, jose benito gonzalez@cern.ch](#) [s. iannidis](#)

Worldwide

This record is related to at least one identifier schema

<

European Open Science Cloud (EOSC)



[Share](#)

OSF

Open Science Framework

[Data Source](#) » *Data Repository* • Compatibility: collected from a compatible aggregator • OpenAIRE Text Mining

Web page: <https://osf.io/> →

DOI: [10.25504/fairsharing.g4z879](https://doi.org/10.25504/fairsharing.g4z879)

re3data: [r3d100011137](https://re3data.org/r3d100011137)

Summary

Publications (126,920)

Research data (751)

Research software (97)

Other research products (291)

Related Data sources

[Statistics](#)

Collected full-texts

122

Countries

United States

Subjects

Humanities and Social Sciences, Life Sciences, Natural Sciences, Engineering Sciences, Social and Behavioural Sciences, Economics, Biology, Medicine, Microbiology, Virology and Immunology, Agriculture, Forestry, Horticulture and Veterinary Medicine, Chemistry, Physics, Geosciences (including Geography), Computer Science, Electrical and System Engineering, Agriculture, Forestry, Horticulture and Veterinary Medicine

Detailed information @

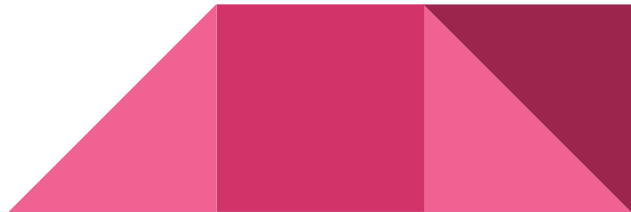
re3data.org, [FAIRsharing](https://fairsharing.org)

Organizations

COS, Center for Open Science partners

DataCite Repository Finder

- Permite conocer detalles sobre los contenidos de cada repositorio.
- Enfocado en recursos que utilizan DOI como identificador persistentes.
- Mantiene un catálogo propio sobre repositorios (DataCite fábrica) que sigue recomendando el uso de Re3data como fuente primaria de información, pero la complementa.



If you plan to deposit your research data in this repository, go to <https://fdz.dzhw.eu/en/>.

133 Works

Filter Works

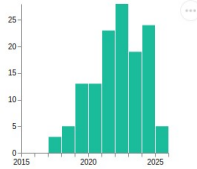
Creators & Contributors ?

☐ Daniel, Andreas	34
☐ İkiz-Akinci, Dilek	16
☐ Multrus, Frank	11
☐ Weber, Anne	6
☐ Heger, Christophe	5
☐ Just, Annika	5
☐ Gross, Christiane	4
☐ Goldan, Lea	4
☐ Euler, Thorsten	4
☐ Fabian, Gregor	4

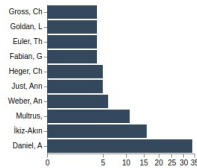
Publication Year

☐ 2025	5
☐ 2024	24
☐ 2023	19
☐ 2022	28
☐ 2021	23
☐ 2020	13
☐ 2019	13
☐ 2018	5
☐ 2017	3

Publication Year



Top Creators and Contributors

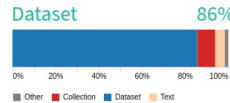


17. Sozialerhebung (2003)

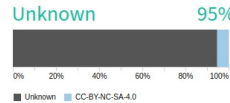
Wolfgang Isserstedt, Elke Middendorff, Steffen Weber, Klaus Schnitzer, Andrä Wolter & Deutsches Zentrum Für Hochschul- Und Wissenschaftsforschung (DZHW)

Version 2.0.0 of Umfrage Und Aggregatdaten Survey And Aggregate Data published 2020 in [Research Data Center for Higher Education Research and Science Studies \(FDZ-DZHW\)](#)

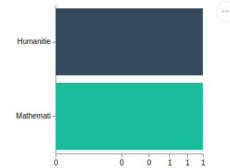
Work Types



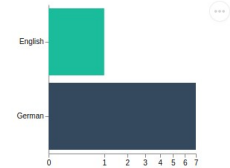
Licenses



Fields of Science



Work Languages



See the repositories in [re3data](#) that meet the FAIR criteria.

Criteria Compliance ?

- ☐ Enabling FAIR Data
- ☐ Project
- ☐ FAIR's FAIR Project

Certificates

- ☐ Core Trust Seal
- ☐ Wds
- ☐ Rat Swd
- ☐ Dsa

Software

- ☐ unknown
- ☐ Fedora
- ☐ Unknown
- ☐ other

Education Works People Organizations Repositories Pages Support Sign In

9 Repositories

Research Data Center for Higher Education Research and Science Studies (FDZ-DZHW)

The Research Data Center for Higher Education Research and Science Studies (FDZ-DZHW) at the German Centre for Higher Education Research and Science Studies (DZHW) in Hannover provides the scientific community with quantitative and qualitative research data from the field of higher education and science studies for research and teaching purposes. The data pool of the Research Data Centre is based on two sources: Firstly, it contains the current surveys of the panels conducted in-house (especially DZHW Graduate Panel, Social Survey, DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification, DZHW Scientists Survey), which are integrated by default. Secondly, the Research Data Centre constantly processes, documents and integrates inventory data of the DZHW and its prior organisations. External data from the research area is also integrated into the FDZ data pool.

humanities and social sciences | humanities and social sciences | education sciences | research on teaching, learning and training | social sciences

empirical social research | social and behavioural sciences | secondary use of data | microdata | higher education studies | university research

family history | survey | educational trajectories | empirical social research

More info about Research Data Center for Higher Education Research and Science Studies (FDZ-DZHW) Repository

Go to Research Data Center for Higher Education Research and Science Studies (FDZ-DZHW) Repository

Research Library for the History of Education at DIPF (BBF)

The BBF | Research Library for the History of Education in Berlin is a research library with outstanding library and archive holdings and a wide range of digital services on German educational history in its international context. It is a center of historical educational research in Germany and is part of the DIPF | Leibniz Institute for Research and Information in Education.

humanities and social sciences | education sciences | social and behavioural sciences | quantitative data | qualitative data | teaching material

pedagogy

More info about Research Library for the History of Education at DIPF (BBF) Repository

Go to Research Library for the History of Education at DIPF (BBF) Repository

Australian Antarctic Data Centre

DataCite Commons

Type to search... Works People Organizations Repositories Pages Support Sign In

Research Data Center for Higher Education Research and Science Studies (FDZ-DZHW)

<https://doi.org/10.17616/r3br0g>

133 Works 22 Citations

The Research Data Center for Higher Education Research and Science Studies (FDZ-DZHW) at the German Centre for Higher Education Research and Science Studies (DZHW) in Hannover provides the scientific community with quantitative and qualitative research data from the field of higher education and science studies for research and teaching purposes. The data pool of the Research Data Centre is based on two sources: Firstly, it contains the current surveys of the panels conducted in-house (especially DZHW Graduate Panel, Social Survey, DZHW Panel Study of School Leavers with a Higher Education Entrance Qualification, DZHW Scientists Survey), which are integrated by default. Secondly, the Research Data Centre constantly processes, documents and integrates inventory data of the DZHW and its prior organisations. External data from the research area is also integrated into the FDZ data pool.

Data Access	restricted
	open
Persistent Identifier	doi
Certificates	RatSWD
Data Upload	restricted
Provider Type	dataProvider

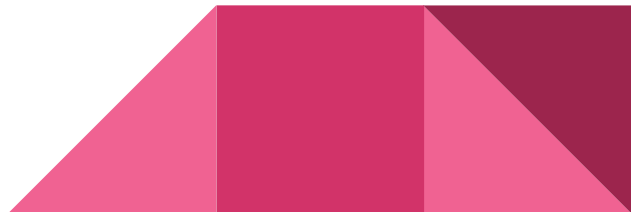
humanities and social sciences | humanities and social sciences | education sciences | research on teaching, learning and training | social sciences

empirical social research | social and behavioural sciences | secondary use of data | microdata | higher education studies | university research

family history | survey | educational trajectories | empirical social research

Conclusiones

- Ninguno de los catálogos es 100% código abierto.
- Tampoco es transparente el proceso de despliegue y configuración.
- No existen mecanismos de descubrimiento de repositorios
 - FAIRICAT?
- En el área de educación, se constatan carencias tecnológicas
 - Basado fuertemente en el enfoque cualitativo.



Dudas, comentarios y discusión

...

¡Gracias!

Gonzalo Torterolo
gonzatorte@gmail.com

