

Generalist Repository Ecosystem Initiative (GREI) to support NIH data sharing and discovery

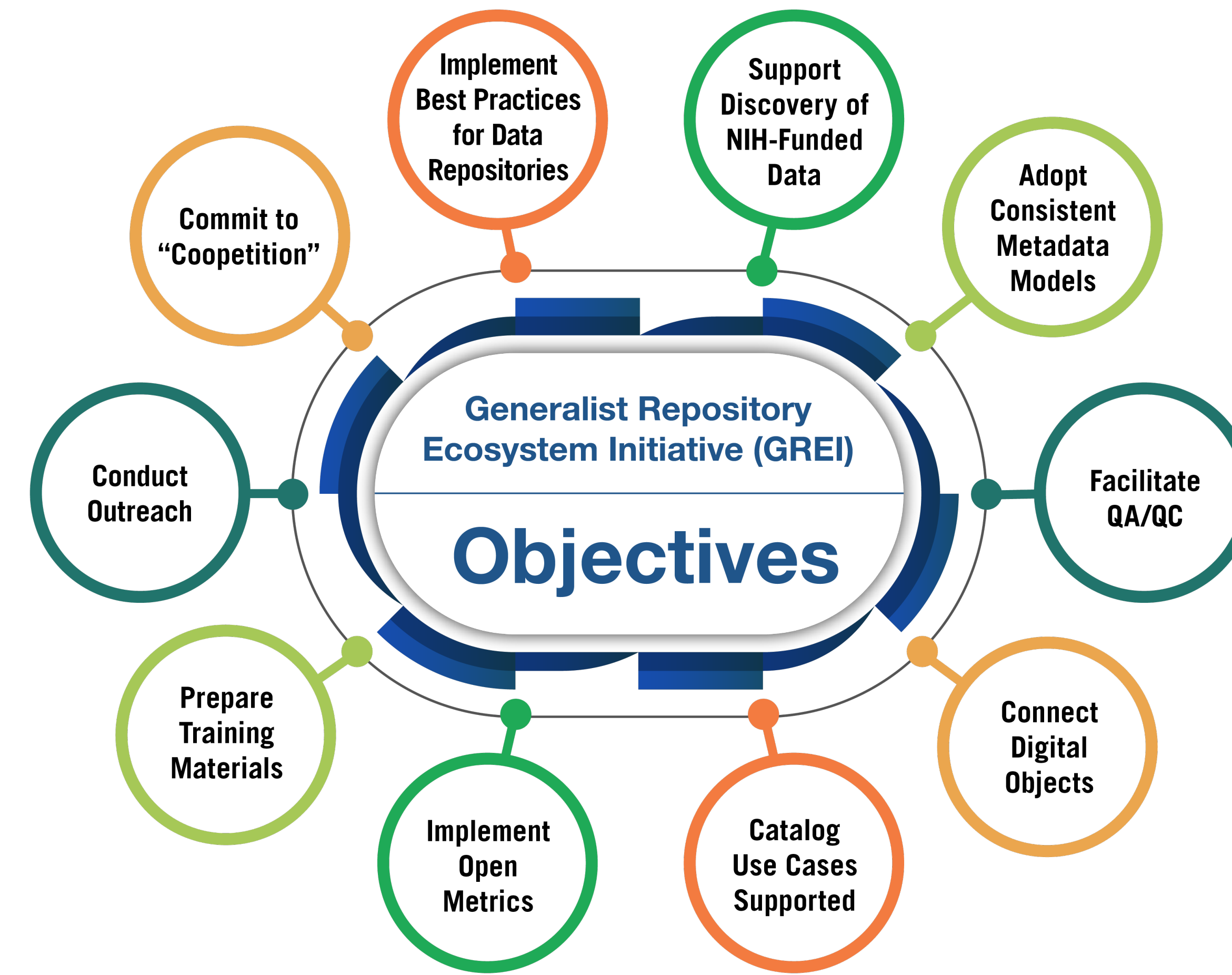
Ana Van Gulick¹,
Kristi Holmes^{2,3},
Sara Gonzales^{2,3},
Julie Wood⁴,
Pearl Go^{2,3}



Figshare¹, Zenodo², Northwestern University³, Vivli⁴

Background

- The NIH announced its new Data Management and Sharing Policy aimed at making research data publicly available in 2020. The policy was implemented in 2023.
- The NIH Office of Data Science Strategy launched the Generalist Repository Ecosystem Initiative (GREI) in February 2022, recognizing the key role that generalist repositories play in the NIH data sharing landscape to support the FAIR sharing of data in trusted repositories.
- GREI brings together seven established generalist repositories to enhance support for NIH data sharing and discovery by
 - Establishing a common set of cohesive and consistent GR capabilities, services, metrics, and social infrastructure, and
 - Providing needed support to librarians and data managers assisting with data sharing in GRs.
- GREI repositories are working together in “coopetition” (cooperation + competition) to advance repository functionality over the course of the multi-year program.
- Learn more at: bit.ly/ODSSGREI



GREI resources for librarians

Use cases for researchers, institutions, and funders

The GREI repositories have published a use cases catalog, illustrating how generalist repositories fit into the NIH data sharing landscape alongside discipline-specific repositories, giving examples of the capabilities of the GREI repositories.



- Researcher** – flexibility to share data and other research outputs FAIR-ly
- Researcher** - find research data to reuse and build upon in their field



- Institution** - report on open data from an institution for compliance and impact
- Funder** - track published data by funding source and measure impact and reuse

Metadata recommendations

- In Year 1, the GREI repositories determined an agreed upon standard, the DataCite Metadata Schema 4.4.
- In Year 2, the GREI repositories focused on use cases for data sharing and searching to identify and recommend specific metadata fields beyond the DataCite required properties to meet the needs of those use cases:

DOI: [10.5281/zenodo.8101957](https://doi.org/10.5281/zenodo.8101957)

- We hope this common metadata schema will be useful beyond GREI to improve interoperability across data repositories and across the NIH data landscape.



The GREI repositories register DOIs and associated metadata through DataCite, a global community focused on ensuring research outputs and resources are openly available and connected so that their reuse can advance knowledge.



GREI's commitment to 'coopetition'

Repositories *compete* on unique features

VALUE LINE

Repositories *cooperate* on common features & standards

GREI repositories are similar. They all support:

- FAIR data sharing across disciplines
- Strive to adhere to repository best practices
- Leverage community standards such as DataCite metadata and persistent identifiers like ORCID and ROR

They also differ:

- Nonprofit and commercial organizations
- Repositories built with open source and proprietary infrastructures
- Offer varying features such as data visualization, file types and sizes, curation, and controlled access

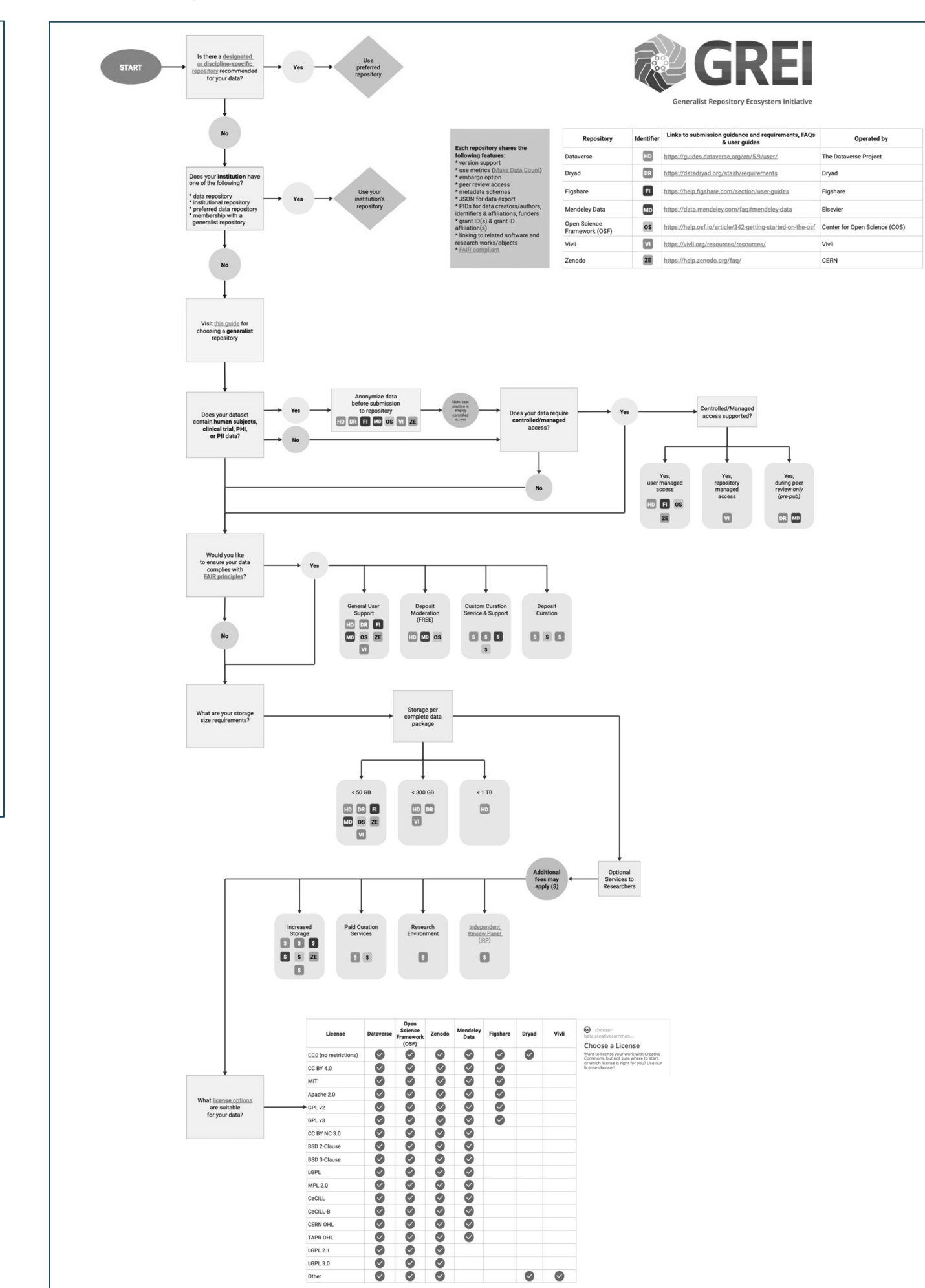
Generalist repository comparison chart and repository selection flowchart

Generalist Repository Comparison Chart
doi:10.5281/zenodo.3946719 | Version 3, 12 May 2023
<https://fairsharing.org/collection/GeneralistRepositoryComparison>

TOPIC	HARVARD DATAVVERSE REPOSITORY	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO	
Brief Description	Harvard Dataverse Repository is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data.	Dryad is an open data archiving platform community committed to the open availability and uniform reuse of data across all disciplines. Dryad facilitates the sharing of data and metadata and provides a Creative Commons Public Domain License (CC0).	Figshare is a free, open access platform for all researchers to share their research data. Dryad facilitates the sharing of data and metadata and provides a Creative Commons Public Domain License (CC0).	Mendeley Data is a free, open access platform for all researchers to share their research data. Dryad facilitates the sharing of data and metadata and provides a Creative Commons Public Domain License (CC0).	OSF is a free and open access platform for all researchers to share their research data. Dryad facilitates the sharing of data and metadata and provides a Creative Commons Public Domain License (CC0).	Well is an independent, non-profit organization that has developed a global data sharing platform for researchers. Well provides a Creative Commons Public Domain License (CC0).	Powering Open Science, built on Open Science, built by researchers for researchers. Well provides a Creative Commons Public Domain License (CC0).	Powering Open Science, built on Open Science, built by researchers for researchers. Well provides a Creative Commons Public Domain License (CC0).
Size Limits	No file size limit per dataset. Harvard Dataverse Repository currently has a file size limit of 2.5GB.	300GB per dataset through massive automation system and to 1TB with assistance from helpdesk/awards.	20GB for free. Figshare offers storage up to 100GB up to 1TB per dataset. System limit of 5TB file.	30GB per dataset	Projects and childhub projects currently have a 50GB storage limit if they are public, and 5GB limit if they are private. There is a 5GB limit for native OSF Storage. There is no limit imposed by OSF for the amount of data sets connected to a given project.	If more than 1TB of study data, reach out to our support@well.org so we can help transfer your data.	50GB per dataset, contact an institution administrator for larger limits.	
Storage space per researcher	1TB	No limit	No limit	No limit	No limit	No limit	No limit	
Persistent, Unique Identifier Support	DOI	DOI	DOI	DOI	DOI	DOI	DOI	

DOI: [10.5281/zenodo.7946938](https://doi.org/10.5281/zenodo.7946938)

The GREI repositories have published a detailed comparison chart breaking down the different features of each GREI repository and a more streamlined flowchart for selecting a repository. Together these resources are meant to support researchers and librarians in determining which generalist repository is the best fit for their data.



DOI: [10.5281/zenodo.11105430](https://doi.org/10.5281/zenodo.11105430)

Training resources

In addition to the resources detailed here, the GREI Zenodo community is home to many other GREI outputs including a library of recordings and slides from GREI webinars, workshops, and other events on topics including:

- Best practices for data sharing in generalist repositories
- Best practices for finding data in generalist repositories
- Including generalist repositories in data management plans (DMPs)
- Data metrics in generalist repositories

All materials are available under the CC BY 4.0 license and we encourage librarians to re-use GREI content.

If there is a training resource or webinar topic that you would like to see from GREI, please contact us via the feedback QR code at right.

Connect with GREI

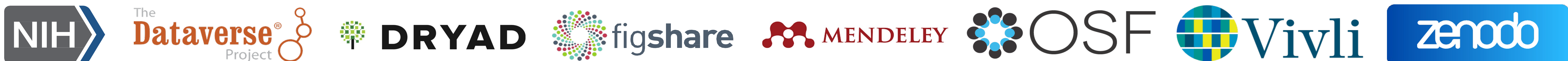
Access GREI resources in our Zenodo community



Share feedback & suggest new training resources or webinar topics



A collaboration between:



Supported by NIH awards OT2DB000001, OT2DB000002, OT2DB000003, OT2DB000004, OT2DB000005, OT2DB000006, OT2DB000013