Bolstering a sustainable generalist repository ecosystem through coopetition

Ana Van Gulick [1], John Chodacki [2, 3], Julie Wood [4], Kristi Holmes [3, 5], Jennifer Gibson [6]
Figshare [1], California Digital Library [2], Zenodo [3], Vivli [4], Northwestern University [5], Dryad [6]

Seven established generalist repositories working together to establish consistent metadata and metrics, develop use cases for data sharing, train and educate researchers on the importance of FAIR data sharing, and more.

ABSTRACT

In February 2022, the NIH Office of Data Science Strategy launched the Generalist Repository Ecosystem Initiative (GREI), which brings together seven generalist repositories (GRs) (Dataverse, Dryad, Figshare, Mendeley Data, Open Science Framework, Vivli, and Zenodo) to enhance support for NIH data sharing and discovery. This program recognizes that GRs play a key role in the NIH data sharing landscape to support the FAIR sharing of data and other research outputs in a trusted repository when they cannot be deposited in a discipline-specific repository.

Key GREI Objectives

- **"Coopetition"** (cooperation + competition) - Collaboration among the GRs to advance repository functionality including through common standards and consistent capabilities.
- **Common metadata** for search and discovery - Creation of a common metadata recommendation on DataCite metadata fields along with recommended controlled vocabularies and persistent identifiers (PIDs) that will advance search and discovery of data across repositories.
- **Common metrics** of reuse - Worked together to define common metrics of data use and have published recommendations for data citations in repositories to demonstrate (re)use and impact of shared data.
- **Data QA/QC** and curation - Curation poses a challenge for standards and scalability as GRs accept data across all disciplines and formats. GREI is currently planning a review of QA/QC practices across repositories to understand opportunities for improvement and define good, better, and best data QA/QC practices for GRs.

The first two years of the GREI coopetition have focused on *reducing the barriers to FAIR data sharing and discovery of data within and across repositories.* Through this, GREI hopes to make GRs a more valuable and sustainable part of the NIH data landscape and plans to work with the NIH community including discipline-specific repositories to continue this work.

DataCite & GREI

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.

Through alignment with the DataCite metadata schema, the GREI repositories register consistent metadata, enabling connectivity of datasets with other digital objects such as articles, researchers, research organizations, grants and funders. In collaboration with the Make Data Count initiative, a set of open metrics are being developed to enable the evaluation of research data reuse and impact.

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

GREI OBJECTIVES AND PROGRESS

GREI Objectives	Progress Years 1 & 2
Align with Desirable Characteristics for Data Repositories	All Repositories published how they meet Desirable characteristics
Implement browse & search for NIH funded data	Google-Style Search and Browse Complete
Develop consistent metadata models	Agreed to common metadata recommendations
Use case support	Key use cases developed
Implement open metrics	More than 50% have implemented Make Data Count

GREI INITIATIVES

Data citation recommendations *to support consistent practices to collect, expose and aggregate citations to open data:*
A consistent approach to data citations is an important step to drive meaningful metrics that provide visibility on data usage, signal the added value of data repositories and enable reporting on the reach of NIH-funded research data. In collaboration with Make Data Count, the GREI repositories have developed a common resource on best practices for handling data citations at repositories: https://doi.org/10.5281/zenodo.10562429.

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

Institution - report on open data from an institution for compliance and impact

Researcher - find research data to reuse and build upon in their field

Funder - track published data by funding source and measure impact and reuse

Metadata recommendations:
One goal of GREI is to support interoperability and discovery of datasets across repositories by establishing common metadata standards for generalist repositories. Having first determined an agreed upon standard, the DataCite Metadata Schema 4.4, focusing on the use cases for data sharing and searching allowed the group to move forward to identify specific metadata fields beyond the DataCite required properties metadata that would meet the needs of those use cases. We also hope this common metadata schema will be useful for data repositories beyond GREI to improve interoperability across data repositories and across the NIH data landscape.

QA/QC PLANNED ACTIVITIES

Planned activities for Year 3 of GREI include:

- A comprehensive review of how QA/QC on data is conducted at each repository.
- A Comprehensive review on approaches to handling personal or sensitive data.

GREI Blog | **Calendar of Events** | **GREI Zenodo Community**

A collaboration between:

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