

Coopetition as a Catalyst for Researcher Engagement with Open Data

ISMB/ECCB 2025
July 20-25, 2025



GREI



DRYAD



figshare



OSF



MENDELEY DATA



Vivli



Agenda

- About GREI
 - Mission
 - Coopetition model
 - Accomplishments
- Common metrics
 - Uniform tracking of impact of open datasets
 - Dataset citations - reward/incentive for researchers
- Connect With GREI



Presenter:

Mark Hahnel
VP of Open Research
m.hahnel@digital-science.com





The
Dataverse
Project



DRYAD

 **figshare**

 **OSF**



MENDELEY DATA



Vivli

 **zenodo**

The Generalist Repository Ecosystem Initiative is an NIH Office of Data Science Strategy program which aims to develop collaborative approaches for data management and sharing through the inclusion of generalist repositories in the NIH data ecosystem, and to better enable sharing, discovery, and reuse of NIH-funded data in generalist repositories.

The GREI program employs a coopetition model in which seven established generalist repositories are working together to fulfill both elements of the GREI mission:

To establish a common set of
cohesive and consistent
capabilities, services, metrics,
and **social infrastructure** across
generalist repositories

To raise general awareness and
help researchers adopt **FAIR**
principles to better share and
reuse data



bit.ly/ODSSGREI

NIH Awards:

3OT2DB000001-01S1
3OT2DB000002-01S1
3OT2DB000003-01S1
3OT2DB000004-01S1
3OT2DB000005-01S1
3OT2DB000006-01S1
3OT2DB000013-01S1

<https://datascience.nih.gov/data-ecosystem/generalist-repository-ecosystem-initiative>

GREI Repositories

Similarities

- FAIR data sharing across disciplines



Findable



Accessible



Interoperable



Reusable

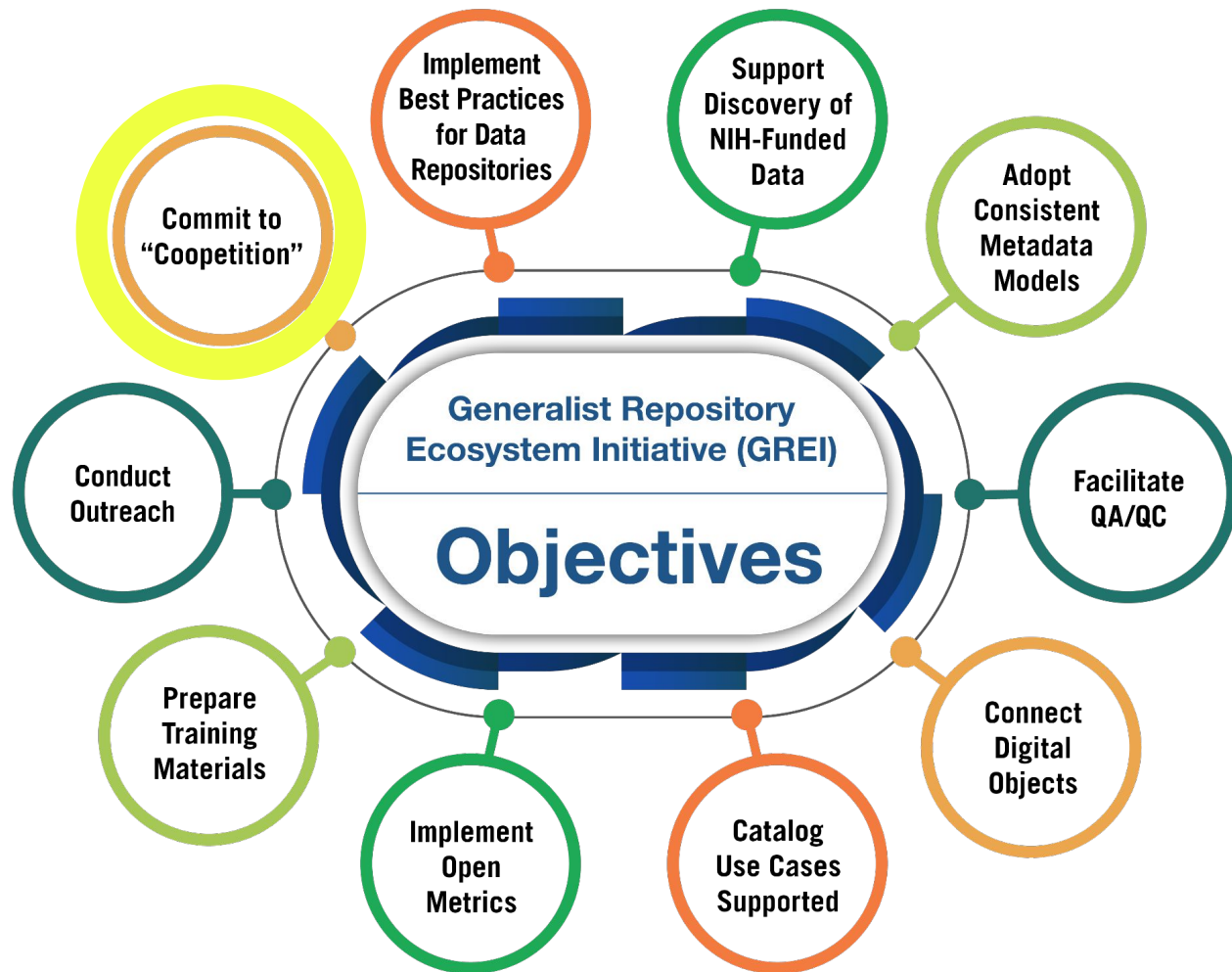
- Adhere to repository best practices
- Leverage community metadata standards and PIDs



Distinct Attributes

- Nonprofit, academic, and commercial organizations
- Built with open source and proprietary infrastructures
- Offer varying features such as:
 - Data visualization
 - File types and sizes
 - Curation
 - Licenses
 - Controlled access







The value of the GREI COOPETITION

Cooperation + Competition = Coopetition

Compete on unique features

— VALUE LINE —

*Cooperate on common features,
standards, and activities
(e.g., metadata, PIDs, metrics, discovery)*

GREI coopetition efficiencies

- **Governance:** unified policies, shared decision-making, common reporting metrics
- **Public accountability:** transparency, coordinated communication, community engagement
- **Innovation:** faster innovation cycles through knowledge sharing, commitment to FAIR principles
- **Operations:** streamlined workflows, avoidance of redundant and cumbersome infrastructures, CQI cycles
- **Products:** metadata, training, guidance, metrics, etc.

System-level synergies through coopetition

- **Effective alternative to a traditional coordinating center**, supporting logistics, governance, collaboration, and communication.
- **Efficient allocation of time and resources** among participating organizations.
- Development of **system-level synergies** that exceed what each could accomplish alone.
- Cultivation of a **partnership-oriented mindset** among traditionally competing groups



DRYAD



figshare



OSF



MENDELEY DATA



Vivli



GREI Results → Enhanced Data Sharing and Discovery

Enhanced generalist repository **functionality** for data sharing and discovery use cases that lower the barriers to sharing and reusing open data.

- GREI repositories have implemented **Google-style search and browse**, and most have further implemented **advanced search** and **search results filtering and sorting** functions. Work to fully index GREI repositories for search in existing cross-repository search tools is ongoing.
- GREI repositories developed a recommendation for which optional **DataCite metadata fields** repositories should implement for consistency and all are in the process of implementing.
- All GREI repositories have implemented or are working to implement **Make Data Count standards** and the **DataCite usage tracker**.
- Almost all GREI repositories **collect and submit relational metadata to DataCite**.
- All GREI repositories have implemented or are working to **implement ROR IDs** for author affiliations and funder organizations.




Implications for the Larger Data Sharing Landscape

A stronger repository landscape that provides researchers with **trusted generalist repositories** to catalyze the flexible sharing of data and other research outputs in any format that cannot be shared elsewhere.



Recommendations for **common repository standards** for metadata, persistent identifiers, ontologies, metrics, and citations, which are openly available for adoption by other repositories to support broad interoperability across repositories.

The importance of data citations has been recognized by earlier community work, for example, in the form of the [FORCE11 Data Citation Principles](#), which cover the purpose, function and attributes of citations to data. The principles note as part of its [guidance](#) that while citations may vary in style, citations to data "should be included in the full reference list along with



Generalist Repository Ecosystem Initiative

GREI Data citation best practices for repositories

Introduction

One of the objectives of the [Generalist Repository Ecosystem Initiative](#) (GREI) is to implement data metrics that enable reporting on the reach and impact of NIH-funded research data.

Data citations are a key component of the measures of data usage, as they bring benefits to the data creators, the data users, and the scholarly communication ecosystem more broadly:

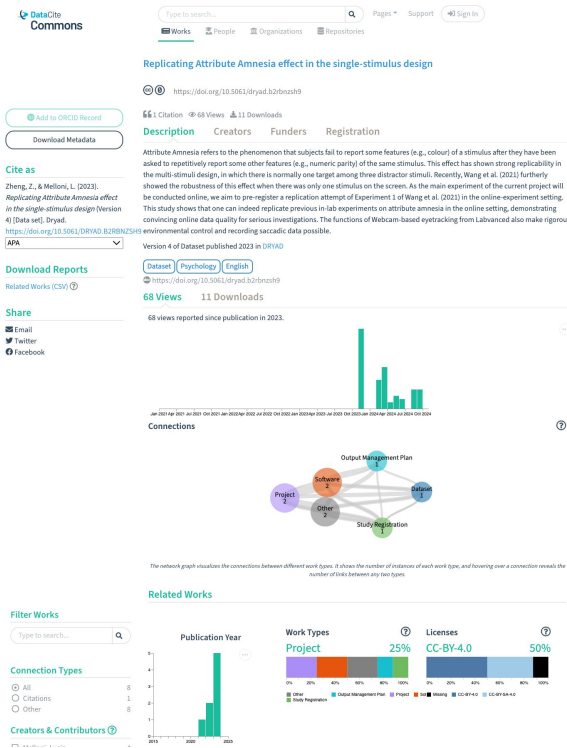
- Data citations are a signal of a dataset being used in research (beyond mere exploration), providing valuable information to evaluate data usage.
- Data citations provide credit for the data producer, the citation recognizes the individual(s) or organization(s) that collected and shared the data used in the citing work.
- For academic researchers, accruing citations to datasets can also be valuable as part of research evaluation frameworks (e.g. for hiring or promotion), as they provide evidence of the reach of their open datasets.
- Surveys of researchers regularly show that getting citations to their research papers as well as citations to the datasets themselves are among the biggest motivators for them



Explore DataCite Metadata Store



DOI insights



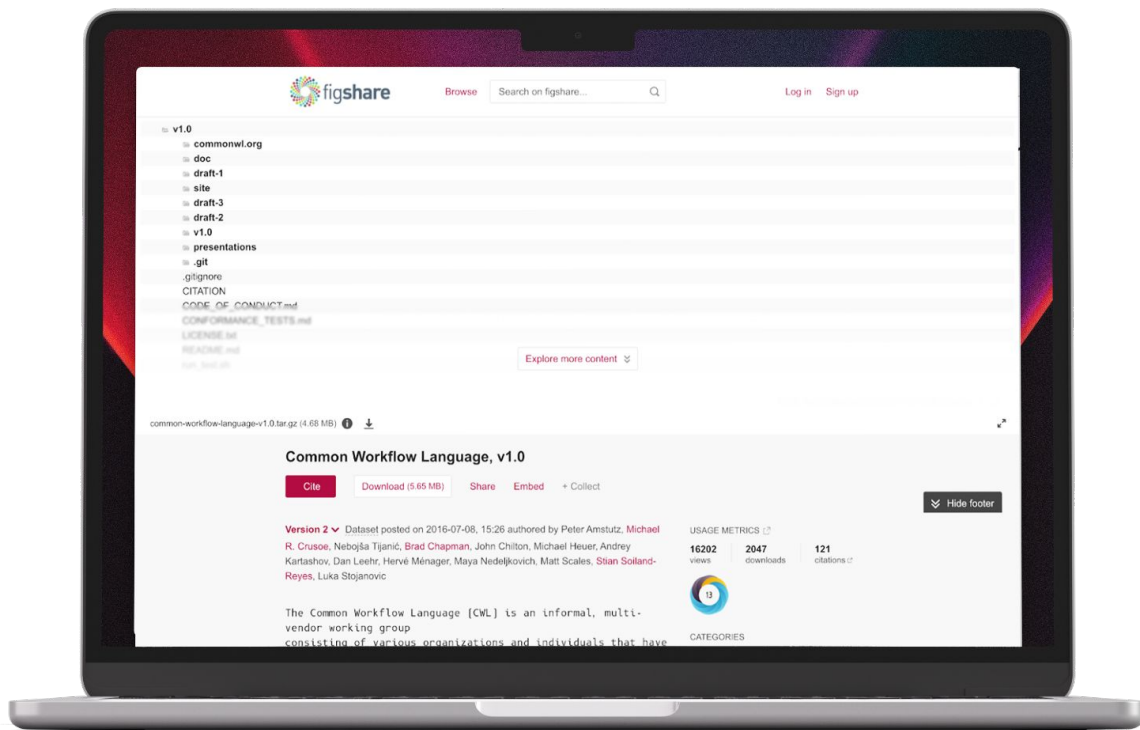
Repository insights



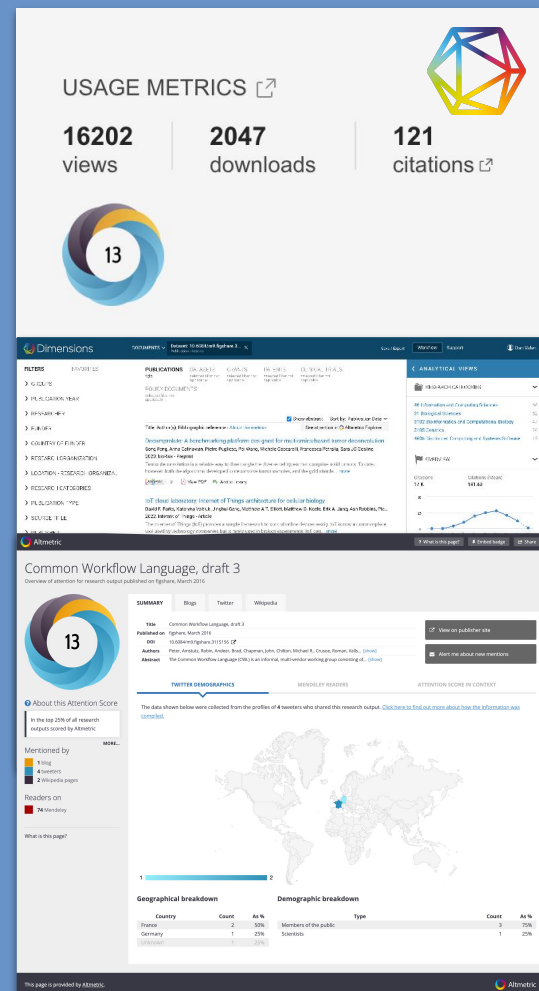
<https://commons.datacite.org/>



Figshare | Track attention and use/reuse



<https://doi.org/10.6084/m9.figshare.3115156.v2>



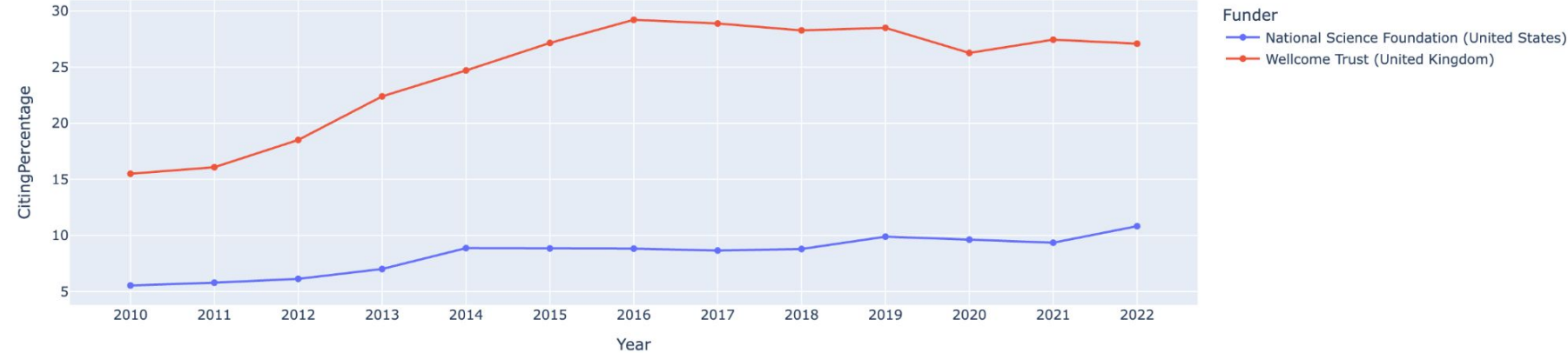
Citing Percentage Over Time (2010–2022)

Funder × ▼

× National Science Foundation (United States) × Wellcome Trust (United Kingdom) × ▼



Citing Percentage Over Time for Selected Funders



Citing Percentage Over Time (2010–2022)

Affiliation × ▼

× University of Liverpool (United Kingdom) × Liverpool John Moores University (United Kingdom)

× Edge Hill University (United Kingdom) |

Citing Percentage Over Time for Selected Affiliations





NIH Data Sharing Index (S-index) Challenge

Promoting data sharing and developing a robust metric to reward exemplary data sharers.

This Challenge aims to incentivize and reward data sharing excellence, promoting a new metric for assessing how effectively researchers share valuable data, driving a culture of openness in science.



Apply starting 09/17/25

 Follow challenge (122)

 Share

Submission period: Phase 1 closed / Phase 2 opens on 09/17/25 01:00 PM GMT+1

Challenge type: Scientific

Total cash prizes: \$1,000,000

 [Print challenge](#)



*Access all GREI resources via the NIH ODSS
GREI landing page → bit.ly/ODSSGREI*



*Share your feedback on GREI resources,
progress, and future plans via our GitHub
Discussion Board → bit.ly/GREIdiscuss*



 **GREI**
Thank you!