



Engaging Research Infrastructure Providers in the Transformation to Open Science

*FORCE2024
UCLA Campus - August 1, 2024*



Meet the panelists



Kristi Holmes

Associate Dean for
Knowledge Management and
Strategy &, Feinberg School
of Medicine, Northwestern
University

GREI Coopetition Year 3
Co-chair



Sarah Sweet

Senior Clinical Research
Manager, Vivli



Lisa Federer

Acting Director, Office
of Strategic Initiatives,
National Library of
Medicine

GREI Program Team



Ana Van Gulick

Government &
Funder Lead, Figshare

GREI Coopetition
Year 2 Co-chair



Panel Session Agenda

- The “Coopetition” Model of Collaboration (Kristi)
- The NIH Perspective (Lisa)
- GREI Recommendations and Resources (Ana)
- Generalist Repository Use Cases (Sarah)
- Implications for the Larger Data Sharing Landscape (Kristi)



The “Coopetition” Model of Collaboration

Kristi Holmes, Northwestern University & Zenodo





NIH Generalist Repository Ecosystem Initiative

The mission of GREI is to establish a common set of capabilities, services, metrics, and social infrastructure; raise general awareness and facilitate researchers to adopt FAIR principles to better share and reuse data.

This initiative will further enhance the biomedical data ecosystem and help researchers find and share data from NIH-funded studies in generalist repositories.

NIH Awards:
3OT2DB000001-01S1
3OT2DB000002-01S1
3OT2DB000003-01S1
3OT2DB000004-01S1
3OT2DB000005-01S1
3OT2DB000006-01S1
3OT2DB000013-01S1



Goals of the Generalist Repository Ecosystem Initiative

- 

1
Make it easier for researchers to **share data**
- 

2
Enable the improved **discoverability** of NIH-funded data across generalist repositories
- 

3
Support greater **reproducibility** of NIH-funded research by ensuring data associated with publications is readily available
- 

4
Avoid **duplication** of data across repositories
- 

5
Encourage NIH-funded researchers to be both contributors and consumers to **increase the reuse of data**

<https://doi.org/10.6084/m9.figshare.21318270>

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





Coopetition = Cooperation + Competition

Coopetition

- Originally coined by Brandenburger & Nalebuff, 1996
- Sayeed Choudhury keynote at February 2020 NIH Workshop on the Role of Generalist and Institutional Repositories to Enhance Data Discoverability and Reuse

Coopetition



Compete on Unique Features

VALUE LINE

Cooperate on common features and standards

(metadata, PIDs, metrics, discovery, privacy)



GREI coopetition = *collaboration among the generalist repositories to jointly advance repository functionality to support NIH data sharing, discovery, and reuse*



Coopetition allows for:

- Development of a **cohesive and interoperable** generalist repository landscape
- Regular **communication** across repositories
- Implementation of **common best practices and standards**
- Leveraging existing **community standards** (e.g. DataCite, ORCID, ROR)
- Enhanced **flexible data sharing** for NIH funded researchers
- Shared **training and outreach** for generalist repository use cases
- Improved **data discovery and impact** tracking across repositories
- Unified **partnerships** with community stakeholders including institutions and funders
- *Also allows repositories to offer **varying features** such as visualization and analysis, tool integrations, custom metadata, and advanced functionality for specific use cases*
- ***Challenges** of working across different business models, infrastructures and product road maps with globally distributed teams*



The NIH Perspective

Lisa Federer, National Library of Medicine, U.S. National Institutes of Health



Focus on Open Science

Work includes:

- Updating or developing plans for increasing public access to federally funded research
- Funding opportunities
- Infrastructure enhancements
- Training and capacity building
- Broadening participation in open science
- Community engagement



Open Science Announcements from Federal Agencies

Open Science is the principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity.

Federal agencies are celebrating 2023 as a Year of Open Science, a multi-agency initiative across the federal government to spur change and inspire open science engagement through events and activities that will advance adoption of open, equitable, and secure science.

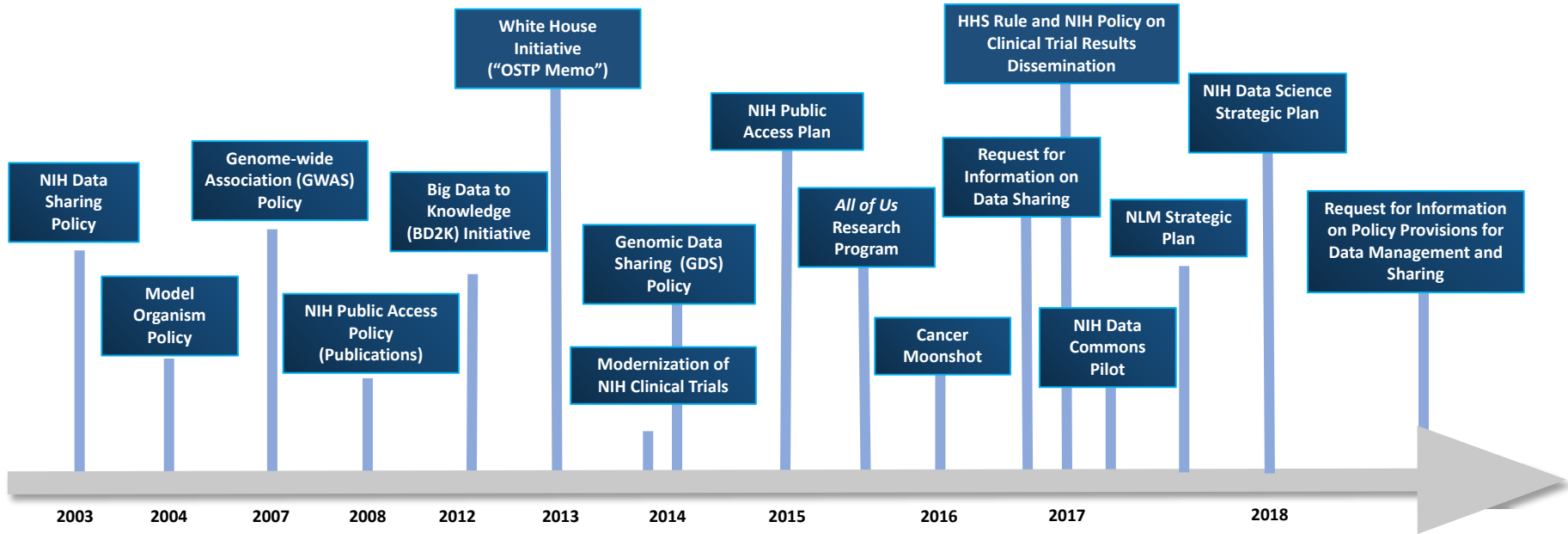
Help spotlight the value of open science by checking out this [toolkit](#): choose and use images for social media, presentations, posters and virtual backgrounds—agency partners may even co-brand. #YearOfOpenScience.

Centers for Disease Control and Prevention
Department of Agriculture
Department of Commerce
Department of Energy
Department of State
Department of Transportation
Environmental Protection Agency
National Aeronautics and Space Administration
National Endowment for the Humanities

National Institutes of Health
National Institute of Standards and Technology
National Oceanic and Atmospheric Administration
National Science Foundation
Smithsonian Institution
U.S. Geological Survey
U.S. General Services Administration
White House Office of Science and Technology Policy



NIH's Culture of Data Sharing





NIH Policy for Data Management and Sharing (DMS)

- **Submission of Data Management & Sharing Plan for all NIH-funded research**
- **Compliance with the ICO-approved Plan** (*may affect future funding*)
- **Effective January 25, 2023** (*replaces 2003 Data Sharing Policy*)

Selecting a repository: the NIH data sharing landscape

NIH strongly encourages
subject-specific, open access Data Sharing Repositories
as a first choice.

https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

Datasets up to **2 gigabytes**

PubMed Central

Stores publication-related supplemental materials and datasets directly associated publications.



Datasets up to **20 gigabytes**

Generalist Repositories

Datasets associated with publications or otherwise and links to PubMed.



High priority datasets, **petabyte-scale**

Cloud Partners (STRIDES Program)

Store and manage large scale, high priority NIH datasets.



GREI Recommendations and Resources

Ana Van Gulick, Figshare



Generalist Repository Comparison Chart

doi:10.5281/zenodo.3946720

This chart is designed to assist researchers in finding a generalist repository should no domain repository be available to preserve their research data. Generalist repositories accept data regardless of data type, format, content, or disciplinary focus. For this chart, we included a repository available to all researchers specific to clinical trials (Vivli) to bring awareness to those in this field.

<https://fairsharing.org/collection/GeneralRepositoryComparison>

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Brief Description	Harvard Dataverse 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.	Open-source, community-led data curation, publishing, and preservation platform for CCO publicly available research data Dryad is an independent non-profit that works directly with: <ul style="list-style-type: none"> researchers to publish datasets utilizing best practices for discovery and reuse publishers to support the integration of data availability statements and data citations into their workflows institutions to enable scalable campus support for research data management best practices at low cost 	A free, open access, data repository where users can make all outputs of their research available in a discoverable, reusable, and citable manner. Users can upload files of any type and are able to share diverse research products including datasets, code, multimedia files, workflows, posters, presentations, and more. With discoverable metadata supporting FAIR principles, file visualizations, and integrations, researchers can make their work more impactful and move research further faster.	Mendeley Data is a free repository specialized for research data. Search more than 20+ million datasets indexed from 1000s of data repositories and collect and share datasets with the research community following the FAIR data principles.	OSF is a free and open source project management tool that supports researchers throughout their entire project lifecycle in open science best practices.	Vivli is an independent, non-profit organization that has developed a global data-sharing and analytics platform. Our focus is on sharing individual participant-level data from completed clinical trials to serve the international research community.	Powering Open Science, built on Open Source. Built by researchers for researchers. Run from the CERN data centre, whose purpose is long term preservation for the High Energy Physics discipline, one of the largest scientific datasets in the world
Size limits	No byte size limit per dataset. Harvard Dataverse currently sets a file size limit of 2.5GB.	300GB/dataset	Soft limit of 20GB/file for free accounts. System limit of 5000GB/file. Unlimited storage of public data but 20GB storage for private data for free accounts. Email info@figshare.com to have upload and storage limits raised.	10GB per dataset	Projects currently have not storage limit. There is a 5GB/file upload limit for native OSF Storage. There is no limit imposed by OSF for the amount of storage used across add-ons connected to a given project.	If more than 10GB per study data, reach out to us	50GB per dataset, contact us via https://zenodo.org/support for higher limits
Storage space per researcher	1 TB per researcher	No limit	No limit	No limit	No limit	No limit	No limit
Persistent, Unique Identifier Support	DOI, Handle	DOI	DOI	DOI	DOI	DOI	DOI

<https://doi.org/10.5281/zenodo.3946719>

Common and Unique Repository Features:

Common:

Core Metadata
 Persistent Identifiers
 Discoverable
 Flexibility
 Open access, FAIR
 Metrics

Unique:

Output types
 Storage, size limits
 Licenses
 Curation
 Controlled Access
 Visualization
 Costs



GREI Metadata Recommendation



By endorsing and adopting a common metadata standard, the GREI repositories support interoperability, discoverability, and transparency in data sharing.

The GREI repositories worked together to identify and recommend the most relevant optional **DataCite Metadata Schema 4.4** fields for data repositories to collect. The recommendation has been published for the benefit of repositories outside of GREI.

The GREI recommendation will continue to evolve based on community feedback, new use cases, and emerging community initiatives.

GREI Metadata and Search Subcommittee: Recommendations from DataCite schema version 4.4

Version 01: Last updated 2023-06-29

Overview

One goal of GREI is to support interoperability and discovery of datasets across repositories by establishing common metadata standards for the generalist repositories. Having focused on an agreed standard, the [DataCite Metadata Schema 4.4](#), the GREI Metadata and Search subcommittee has set its Year 2 goal for repositories to build on their existing work on metadata for research datasets. Focusing on a few high-level use cases for data sharing and searching allowed the group to move forward to identify specific metadata beyond the DataCite required properties metadata that would meet the needs of those use cases.

With the inclusion of DataCite as a GREI stakeholder, more opportunities have been reviewed to now provide a recommendation to the GREI repositories to add additional metadata fields and enhance the quality of the metadata being provided. The subcommittee has continued having detailed discussions ensuring that GREI repositories collect and provide metadata in a way that is useful to all stakeholders.

With this in mind, the GREI Metadata and Search subcommittee has created this recommendation to strongly encourage that each repository member collect the following metadata to support the generalist repository use cases for sharing, discovering and tracking the impact of data.

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.

Recommendation

The document lists strongly encouraged metadata to be collected by each GREI repository in alignment with the metadata collected by DataCite's optional metadata properties. Where applicable, the values and vocabularies that repositories are encouraged to use have also been reviewed by the subcommittee and included in the recommendations.

While these are the recommendations of the subcommittee to-date, the goal of the subcommittee is for the recommended common schema to evolve based on community feedback, new use cases, and emerging community initiatives to allow for updates and additions to the strongly recommended fields in the future.

For DataCite's property definitions, cardinality/quantity constraints, and other usage notes, see [DataCite Metadata Schema Documentation v4.4](#).



Catalog Use Cases Supported (covered in more detail later)

Support discovery for NIH-funded Data

- Focus on similar Use Cases
- Act as exemplar for other repos
- Spark conversations with other stakeholders

Use cases documented so far:

- As an **NIH-funded researcher**, I want to select a repository to share my data...
- As a **researcher**, I want to find research data of interest...
- As an **institution**, I want to report on all datasets from my institution...
- As a **funder**, I want to find datasets we have funded...
- As an **institution**, I want to capture and preserve research data from my institution by using repositories/services...

Use Case documents can be found on Zenodo:

<https://zenodo.org/communities/grei/records?q=&f=subject%3AUse%20Cases&l=list&p=1&s=10&sort=newest>

As an institution, I want to REDORT on all datasets from my institution to the relevant Data Repository so that I can ensure compliance with policies and TRACIS impact management plan commitments by our researchers.

This use case highlights ways funders can leverage generalist repositories to track compliance with data sharing policies and understand data reuse.

Background:
The University's Research Director at a Carnegie Mellon University leads the subject of the research project. The University's Research Director is responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments. The University's Research Director is also responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments.

As a funder from a specific NIH institution of a general, I want to REDORT on all datasets from my institution to the relevant Data Repository so that I can ensure compliance with policies and TRACIS impact management plan commitments by our researchers.

This use case highlights ways funders can leverage generalist repositories to track compliance with data sharing policies and understand data reuse.

Background:
The University's Research Director at a Carnegie Mellon University leads the subject of the research project. The University's Research Director is responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments. The University's Research Director is also responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments.

As an NIH-funded researcher, I want to USE Figshare to share my data, so that I can comply with my data management and sharing plan and the conditions of my grant.

This use case highlights ways researchers can leverage generalist repositories to share project data.

Background:
The University's Research Director at a Carnegie Mellon University leads the subject of the research project. The University's Research Director is responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments. The University's Research Director is also responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments.

As an NIH-funded researcher, I want to USE Mandible Data to share my data, so that I can comply with my data management and sharing plan and the conditions of my grant.

This use case highlights ways researchers can leverage generalist repositories to share project data.

Background:
The University's Research Director at a Carnegie Mellon University leads the subject of the research project. The University's Research Director is responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments. The University's Research Director is also responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments.

As an investigator, I want to FND research data of interest in order to find research data of interest in my field.

This use case highlights ways researchers can leverage generalist repositories to find datasets of interest.

Background:
The University's Research Director at a Carnegie Mellon University leads the subject of the research project. The University's Research Director is responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments. The University's Research Director is also responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments.

As an NIH-funded researcher, I want to FND research data of interest in order to find research data of interest in my field.

This use case highlights ways researchers can leverage generalist repositories to find datasets of interest.

Background:
The University's Research Director at a Carnegie Mellon University leads the subject of the research project. The University's Research Director is responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments. The University's Research Director is also responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments.

As an NIH-funded researcher, I want to REDORT on all datasets from my institution to the relevant Data Repository so that I can ensure compliance with policies and TRACIS impact management plan commitments by our researchers.

This use case highlights ways researchers can leverage generalist repositories to track compliance with data sharing policies and understand data reuse.

Background:
The University's Research Director at a Carnegie Mellon University leads the subject of the research project. The University's Research Director is responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments. The University's Research Director is also responsible for ensuring that the University's research data is shared in a way that is compliant with the University's policies and the TRACIS impact management plan commitments.



GREI Data Citation Best Practices



In partnership with the **MakeDataCount initiative**, the GREI repositories have published data citation best practices for repositories.

This guidance for repositories includes collecting, storing, and exposing citations, as well as a “cite as” template for researchers.

Each of the GREI repositories either already collect data citations or have it on their roadmap. As a group, the repositories are considering approaches to capture those citations in a manner that provides transparent and consistent information to the community on this topic.



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 to publish their data (see for example [The State of Open Data report](#)).
- Data citations increase the rigor and reproducibility of research, enabling data users to document the source of the data they employed as part of research activities.
- Data citations can enable the development of tools to aid search and discovery of research works; increasing visibility on what outputs cite a dataset can help researchers find other scholarly objects relevant to their work.

Providing visibility for data citations is therefore a way of increasing the information on data usage available to the community, it signals the added value of data repositories, and can create an incentive for researchers to share more of their data and to cite open data.

<https://zenodo.org/records/10562429>



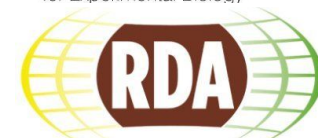
GREI Coopetition Training & Outreach

- 2022 - **GREI Collaborative Webinar Series** on Data Sharing in Generalist Repositories ([10.17605/OSF.IO/JZU37](https://doi.org/10.17605/OSF.IO/JZU37))
- January 2023 - **Generalist Repository Ecosystem Initiative (GREI) Workshop** (sponsored by NIH Office of Data Science Strategy) ([10.5281/zenodo.7714262](https://doi.org/10.5281/zenodo.7714262))
- March 2023 - **RDAP Virtual Summit** - Facilitating use of Generalist Repositories to Share and Discover Data ([10.5281/zenodo.7774200](https://doi.org/10.5281/zenodo.7774200))
- 2023 - **GREI Collaborative Webinar Series**
 - Use Cases in Generalist Repositories and Community Feedback ([10.5281/zenodo.8208834](https://doi.org/10.5281/zenodo.8208834))
 - Metadata Recommendations ([10.5281/zenodo.8356815](https://doi.org/10.5281/zenodo.8356815))
 - Meaningful Metrics ([10.5281/zenodo.10371905](https://doi.org/10.5281/zenodo.10371905))
- September 2023 - **NIH Research Festival** - Data Sharing in Generalist Repositories ([10.5281/zenodo.8357607](https://doi.org/10.5281/zenodo.8357607))
- October 2023 - **SciDataCon** - Generalist repository "coopetition" to enhance data sharing and discovery ([10.5281/zenodo.10037596](https://doi.org/10.5281/zenodo.10037596))
- March 2024 - **NIH Data Repositories and Knowledgebases (DRKB) Network Program Meeting** - Bolstering a sustainable generalist repository ecosystem through coopetition ([10.5281/zenodo.10779520](https://doi.org/10.5281/zenodo.10779520))
- March 2024 - **RDAP Summit** - Generalist Repository Metadata Schema 2.0 – Community Feedback ([10.17605/OSF.IO/UHYGV](https://doi.org/10.17605/OSF.IO/UHYGV))
- March 2024 - **Year of Open Science Culminating Conference** (https://osf.io/meetings/YOS_Conference)
 - Open Science Plumbing: Infrastructure Enabling and Catalyzing Policy Implementation
 - The Coopetition Model of Collaboration in the NIH Generalist Repository Ecosystem Initiative
- May 2024 - **Medical Libraries Association Conference** - Generalist Repository Ecosystem Initiative (GREI) to support NIH data sharing and discovery ([10.5281/zenodo.12210174](https://doi.org/10.5281/zenodo.12210174))
- June 2024 - **Open Repositories** - Empowering Global Progress: GREI Coopetition's Role in Standardizing Transparency, Community, and Sustainability Initiatives ([10.5281/zenodo.12784019](https://doi.org/10.5281/zenodo.12784019))



Training and Community Engagement: Community Partnerships

- Research Data Access and Preservation Association (RDAP)
- FASEB - Salons and 2024 DataWorks! Prize
- RDA and RDA-US - conference sessions
- Network of the National Library of Medicine (NNLM) - Webinars
- ROR - Collaboration on ROR for Repositories
([10.5281/zenodo.10049153](https://doi.org/10.5281/zenodo.10049153))
- Make Data Count - Collaboration on GREI Data citation best practices for repositories ([10.5281/zenodo.10562428](https://doi.org/10.5281/zenodo.10562428))
- NIH Library - Webinars
- DataCite - GREI collaborating member
- The Carpentries - GREI collaborating member



CONNECTING RESEARCH,
ADVANCING KNOWLEDGE



THE
CARPENTRIES



Generalist Repository Use Cases

Sarah Sweet, Vivli



What are key use cases identified by GREI?



As an **NIH-funded researcher**, I want to select a repository to share my data, so that I can comply with my data management and sharing plan and the conditions of my grant.



As a **researcher**, I want to find research data of interest so that I can validate findings, reuse data, and build on work within my discipline.



As an **institution**, I want to report on all datasets from my institution, so that I can ensure compliance of research data sharing and management plan commitments by our researchers.



As a **funder** from a specific NIH institute or in general, I want to find datasets we have funded, so that I can report on compliance with policies, and track impact of research funding and usage of data.



As an **institution**, I want to capture and preserve research data from my institution by using several repositories (including a generalist repository).

As an **institution**, I use a generalist repository infrastructure for my institutional data repository.



Use Case 1:

As an **NIH-funded researcher**, I want to **select a repository** to share my data, so that I can comply with my data management and sharing plan and the conditions of my grant.

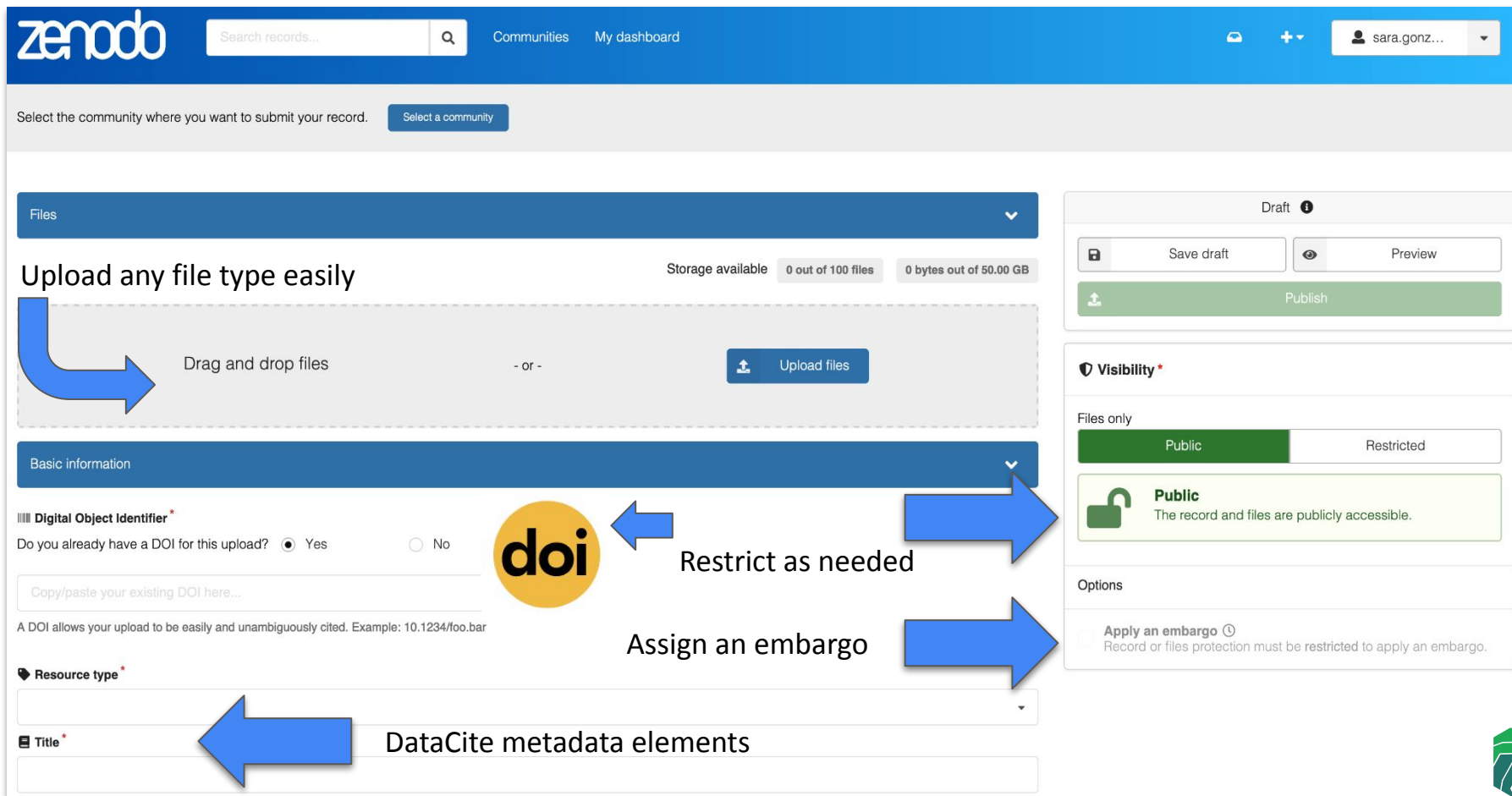
Aspects of the GREI repositories compliant with data sharing mandates:

- Mint DOIs for data deposits, supporting discoverability, citeability, and allowing resource linking
- User-friendly deposit forms with metadata enhancing findability
- Supports upload and/or linking of software/code, supporting reproducibility
- Allow users to assign licenses outlining appropriate data reuse
- Robust metrics to track impact
- Incorporation of PIDs and ontologies, such as ORCID for sign-in, ROR, and CRediT roles
- Support upload of files in any format
- Support controls over public/private aspects of a project or deposit; support embargo
- Dedicated long-term preservation mechanism and/or policy



Use Case 1:

As an **NIH-funded researcher**, I want to **select a repository** to share my data, so that I can comply with my data management and sharing plan and the conditions of my grant.



The screenshot shows the Zenodo upload page. At the top, there is a search bar and navigation links for 'Communities' and 'My dashboard'. Below this, a prompt asks to 'Select the community where you want to submit your record.' with a 'Select a community' button. The main area is titled 'Files' and contains the text 'Upload any file type easily' and 'Storage available 0 out of 100 files 0 bytes out of 50.00 GB'. A large blue arrow points to a dashed box containing 'Drag and drop files' and an 'Upload files' button. Below this is the 'Basic information' section, which includes a 'Digital Object Identifier' field with radio buttons for 'Yes' and 'No', a 'doi' logo, and a text input for 'Copy/paste your existing DOI here...'. A blue arrow points from the 'doi' logo to the 'Restrict as needed' text. Below the DOI field is a note: 'A DOI allows your upload to be easily and unambiguously cited. Example: 10.1234/foo.bar'. The 'Resource type' field is empty, and a blue arrow points from the 'DataCite metadata elements' text to it. The 'Title' field is also empty. On the right side, there is a 'Draft' section with 'Save draft', 'Preview', and 'Publish' buttons. Below that is the 'Visibility' section, which has 'Files only' and 'Public' (selected) and 'Restricted' options. A green box indicates 'Public' with the text 'The record and files are publicly accessible.' Below this is the 'Options' section, which includes an 'Apply an embargo' checkbox and a note: 'Record or files protection must be restricted to apply an embargo.' Two blue arrows point from the 'Restrict as needed' and 'Assign an embargo' text to the 'Apply an embargo' checkbox.



Use Case 2:



As a **researcher**, I want to **find research data of interest** so that I can validate findings, reuse data, and build on work within my discipline.

The GREI repositories' robust search features include:

- Keyword search, advanced search, and state-of-the-art built-in search engines
- Sorting/filtering features: subject, affiliation, resource type, funders, date
- Browse feature: by place name, subject, journal, institution
- Preview of file on record landing page
- Open APIs for downloading record metadata
- Make data findable in search aggregators like Google Dataset Search, DataCite Commons, etc.
- Enable relations between digital objects (e.g., link deposited datasets and publications)
- Enable export of record citations to EndNote and other ref. mgmt tools, and in variety of formats (e.g. JSON)
- PID-enhanced search
- Clear licenses to define what secondary users can do with the data



Use Case 2:



As a researcher, I want to find research data of interest so that I can validate findings, reuse data, and build on work within my discipline.

On figshare.com, users can also search using specific terms and then filter the search results using facets.

The GREI use cases have and will continue to inform the development of Figsare's search and browse capabilities.

The image shows a screenshot of the Figshare search interface with several facets highlighted by orange boxes:

- Content Type:** item (5,993,069), collection (1,419,464), project (5,883)
- Category:** Biological Sciences not else... (2,229,3), Biochemistry (1,932,571), Chemical Sciences not else... (1,893,2), Biotechnology (1,852,065), Medicine (1,847,588)
- Publication date:** Last week, Last month, Last year
- Item Type:** figure (1,980,170), dataset (1,778,805), journal contribution (1,656,698), media (128,541), thesis (105,756)
- Licence:** CC BY 4.0 (3,839,678), CC BY-NC 4.0 (962,808), CC BY + CC0 (399,959), In Copyright (259,668), All Rights Reserved (174,128)
- Funder:** Economic and Social Research ... (16,154), Arts and Humanities Research ... (11,164), European Research Council (7,563), United States Department of Hea... (2,520), Engineering and Physical Scienc... (2,085)
- Source:** publisher (6,323,406), institution (722,357), figshare.com (349,886), preprint (12,284), funder (3,623)
- Publication years:** 2029, 2028, 2027, 2026, 2025, 2024, 2023, 2022, 2021, 2020, 2019, 2018, 2017, 2016, 2015, 2014

Search Figshare content at <https://figshare.com/search/new>



Use Case 3:

As an **institution**, I want to **report on all datasets from my institution**, so that I can ensure compliance of research data sharing and management plan commitments by our researchers.

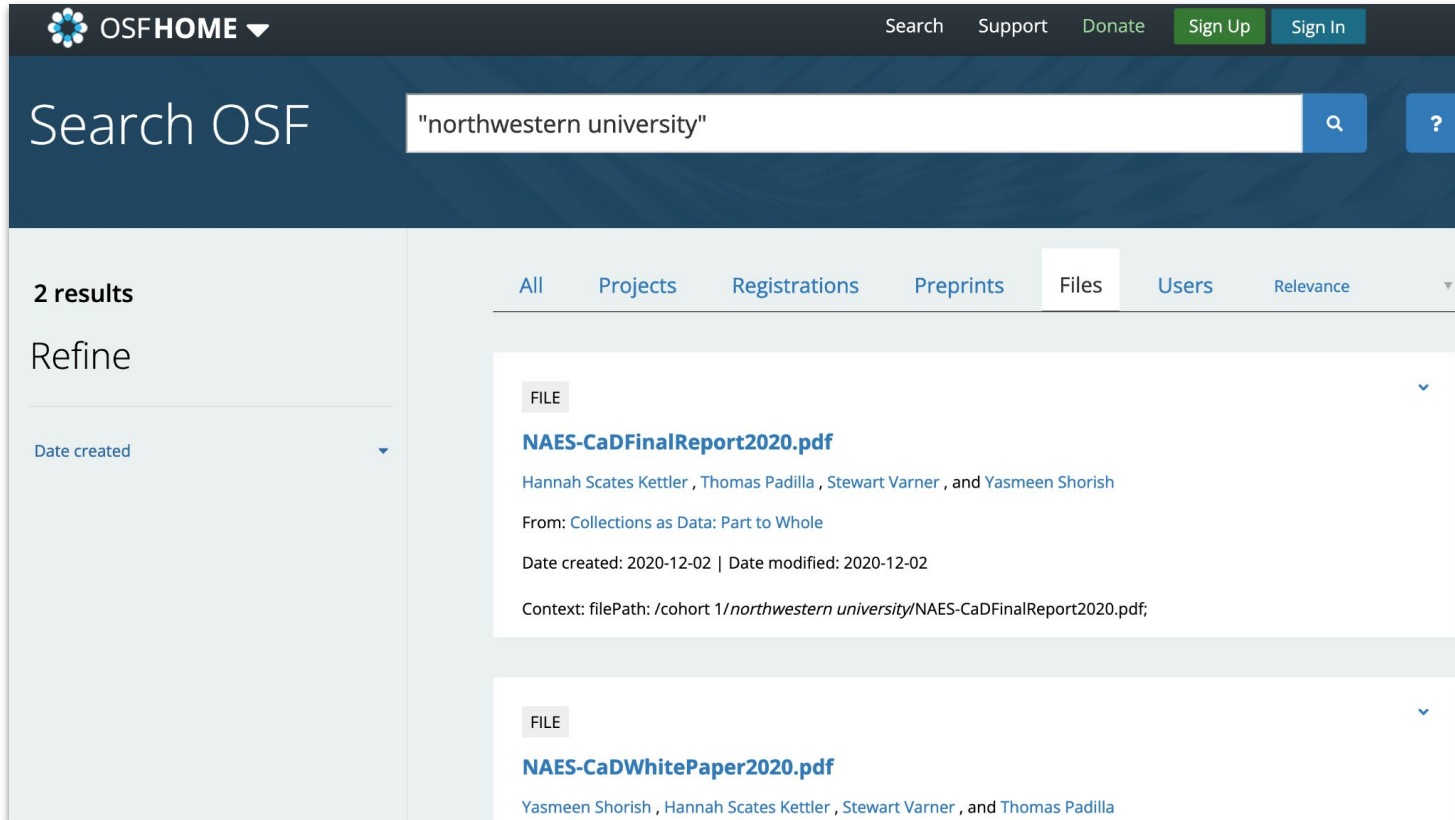
The GREI repositories' features to find data by institution include:

- ❑ Advanced search by institution name, in many cases enhanced by ROR-tagging of institutions
- ❑ Facet/filter searches by institution, and by particular funders
- ❑ Frequent incorporation of popular protocols to pull record metadata such as OAI-PMH, open APIs
- ❑ Administrative dashboards with usage metrics, enabling download of info about shared data as CSV
- ❑ Record metadata shared with DataCite and other aggregators; and available via APIs like CrossRef's



Use Case 3:

As an **institution**, I want to **report on all datasets from my institution**, so that I can ensure compliance of research data sharing and management plan commitments by our researchers.



The screenshot displays the OSFHOME search interface. At the top, the OSFHOME logo is on the left, and navigation links for Search, Support, Donate, Sign Up, and Sign In are on the right. The search bar contains the query "northwestern university". Below the search bar, there are tabs for All, Projects, Registrations, Preprints, Files, Users, and Relevance. The search results are displayed in a list format, showing two results under the "Files" tab. The first result is "NAES-CaDFinalReport2020.pdf" by Hannah Scates Kettler, Thomas Padilla, Stewart Varner, and Yasmeen Shorish. The second result is "NAES-CaDWhitePaper2020.pdf" by Yasmeen Shorish, Hannah Scates Kettler, Stewart Varner, and Thomas Padilla.

OSFHOME

Search Support Donate Sign Up Sign In

Search OSF "northwestern university" ?

2 results

Refine

Date created

All Projects Registrations Preprints Files Users Relevance

FILE

NAES-CaDFinalReport2020.pdf

Hannah Scates Kettler , Thomas Padilla , Stewart Varner , and Yasmeen Shorish

From: Collections as Data: Part to Whole

Date created: 2020-12-02 | Date modified: 2020-12-02

Context: filePath: /cohort 1/northwestern university/NAES-CaDFinalReport2020.pdf;

FILE

NAES-CaDWhitePaper2020.pdf

Yasmeen Shorish , Hannah Scates Kettler , Stewart Varner , and Thomas Padilla



Use Case 4:



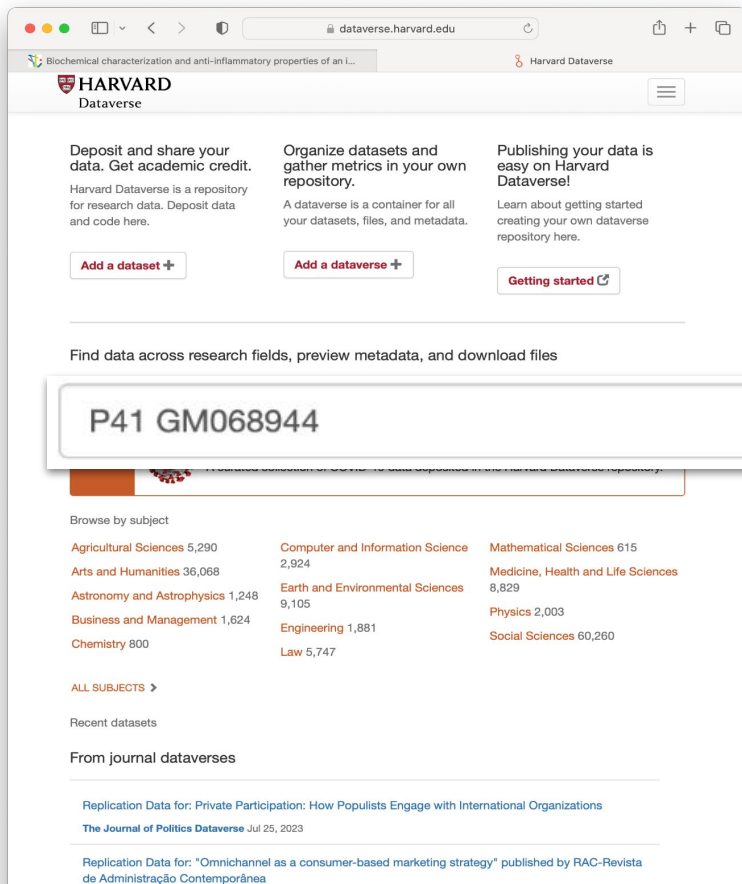
As a **funder** from a specific NIH institute or in general, I want to **find datasets we have funded**, so that I can report on compliance with policies, and track impact of research funding and usage of data.

The GREI repositories' features to find data by funder include:

- ❑ Integration with CrossRef Funder Registry for machine readable data; plus the Funder API lookup
- ❑ Dashboards to see dataset metadata and usage metrics (views, downloads, public disclosures)
- ❑ Altmetric Attention Data and Citation Data of deposited datasets
- ❑ Metadata Fields for funder, grant identifier, award title, award URL; Funder ID included in DOI - and thereby advanced search and filter searches by each of these
- ❑ Sort search results by Grant number
- ❑ PI name and ORCID contained in records confirms data is connected with a particular grant
- ❑ Grant tracking software can track deposited data from the publication, using related identifiers
- ❑ Reach out to the repository's support contact for a report customized by NIH institute



Use Case 4: As a **funder** from a specific NIH institute or in general, I want to **find datasets we have funded**, so that I can report on compliance with policies, and track impact of research funding and usage of data.



Harvard Dataverse


Deposit and share your data. Get academic credit. Harvard Dataverse is a repository for research data. Deposit data and code here.

Organize datasets and gather metrics in your own repository. A dataverse is a container for all your datasets, files, and metadata.

Publishing your data is easy on Harvard Dataverse! Learn about getting started creating your own dataverse repository here.

[Add a dataset +](#) [Add a dataverse +](#) [Getting started](#)

Find data across research fields, preview metadata, and download files

P41 GM068944 

Browse by subject

Agricultural Sciences 5,290	Computer and Information Science 2,924	Mathematical Sciences 615
Arts and Humanities 36,068	Earth and Environmental Sciences 9,105	Medicine, Health and Life Sciences 8,829
Astronomy and Astrophysics 1,248	Engineering 1,881	Physics 2,003
Business and Management 1,624	Law 5,747	Social Sciences 60,260
Chemistry 800		

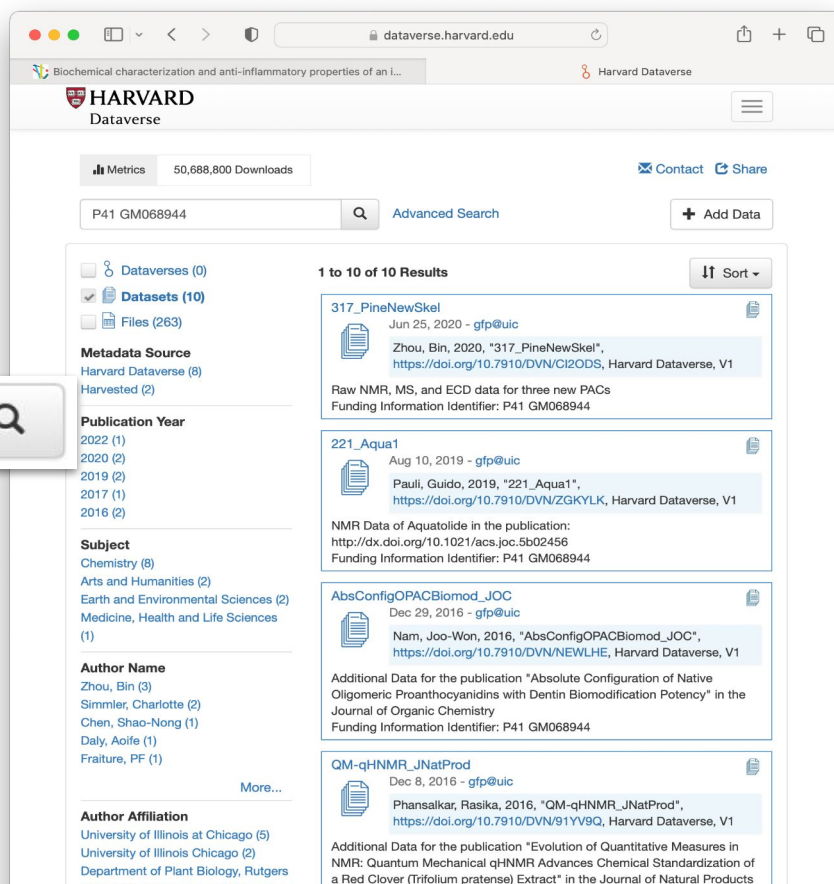
ALL SUBJECTS >

Recent datasets

From journal dataverses

Replication Data for: Private Participation: How Populists Engage with International Organizations
[The Journal of Politics Dataverse](#) Jul 25, 2023

Replication Data for: "Omnicannel as a consumer-based marketing strategy" published by RAC-Revista de Administração Contemporânea



Harvard Dataverse

Metrics 50,688,800 Downloads [Contact](#) [Share](#)

P41 GM068944 [Advanced Search](#) [Add Data](#)

Dataverses (0)
 Datasets (10)
 Files (263)

Metadata Source
Harvard Dataverse (8)
Harvested (2)

Publication Year
2022 (1)
2020 (2)
2019 (2)
2017 (1)
2016 (2)

Subject
Chemistry (8)
Arts and Humanities (2)
Earth and Environmental Sciences (2)
Medicine, Health and Life Sciences (1)

Author Name
Zhou, Bin (3)
Simmier, Charlotte (2)
Chen, Shao-Nong (1)
Daly, Aoife (1)
Fraiture, PF (1)

Author Affiliation
University of Illinois at Chicago (5)
University of Illinois Chicago (2)
Department of Plant Biology, Rutgers University, New Brunswick, New Jer-

1 to 10 of 10 Results [Sort](#)

317_PineNewSkel
Jun 25, 2020 - [gfp@uic](#)
Zhou, Bin, 2020, "317_PineNewSkel", <https://doi.org/10.7910/DVN/Ci2ODS>, Harvard Dataverse, V1
Raw NMR, MS, and ECD data for three new PACs
Funding Information Identifier: P41 GM068944

221_Aqua1
Aug 10, 2019 - [gfp@uic](#)
Pauli, Guido, 2019, "221_Aqua1", <https://doi.org/10.7910/DVN/ZGKYLK>, Harvard Dataverse, V1
NMR Data of Aquatolide in the publication: <http://dx.doi.org/10.1021/acs.jpc.5b02456>
Funding Information Identifier: P41 GM068944

AbsConfigOPACBiomod_JOC
Dec 29, 2016 - [gfp@uic](#)
Nam, Joo-Won, 2016, "AbsConfigOPACBiomod_JOC", <https://doi.org/10.7910/DVN/NEWLHE>, Harvard Dataverse, V1
Additional Data for the publication "Absolute Configuration of Native Oligomeric Proanthocyanidins with Dentin Biomodification Potency" in the Journal of Organic Chemistry
Funding Information Identifier: P41 GM068944

QM-qHNMR_JNatProd
Dec 8, 2016 - [gfp@uic](#)
Phansalkar, Rasika, 2016, "QM-qHNMR_JNatProd", <https://doi.org/10.7910/DVN/91YV9Q>, Harvard Dataverse, V1
Additional Data for the publication "Evolution of Quantitative Measures in NMR: Quantum Mechanical qHNMR Advances Chemical Standardization of a Red Clover (Trifolium pratense) Extract" in the Journal of Natural Products



Use Case 5:

As an **institution**, I want to capture and preserve research data from my institution by using several repositories (including a generalist repository).

As an **institution**, I use a generalist repository infrastructure for my institutional data repository.

Institutional needs include:

- Need a trusted repository that will hold our researchers' outputs long-term, and in all formats
- Need the repository to offer metrics for tracking of our institution's deposits for reporting
- Need the repository to offer PIDs (ORCiDs, ROR) to support credit for our researchers and to enable deposits to be tagged with the consistent spelling of our organization's name
- Need the repository to have robust API for accessing our records




Use Case 5:

As an **institution**, I use a generalist repository infrastructure for my institutional data repository.

Examples of Generalist Repositories being used as Institutional Repositories: UCLA's Dataverse, Carnegie Mellon's Figshare, Northwestern Medicine's use of InvenioRDM/Zenodo



UCLA Dataverse



NRDB Part I: Family Reconstitution
(University of California, Los Angeles)

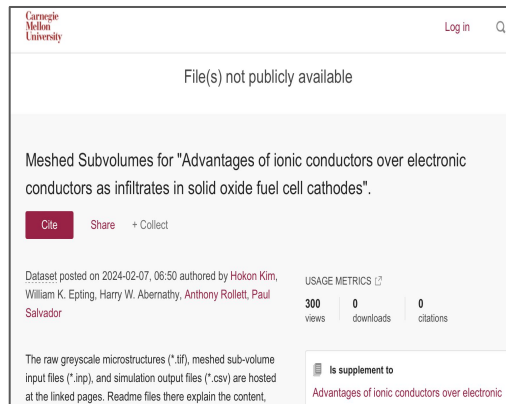
UCLA Dataverse > Neckarhausen Research Database (NRDB) Dataverse > NRDB Part I: Family Reconstitution

Family Reconstitution B: Digital Data New Versions

Version 1.0

 BALL, ROI, 2024, "Family Reconstitution B: Digital Database (Old and New Versions)", <https://doi.org/10.25346/S6/J8VLAJ>, UCLA Dataverse, V1

Cite Dataset ▾ Learn about [Data Citation Standards](#).




Carnegie Mellon University Log in

File(s) not publicly available

Meshed Subvolumes for "Advantages of ionic conductors over electronic conductors as infiltrates in solid oxide fuel cell cathodes".


Cite Share + Collect

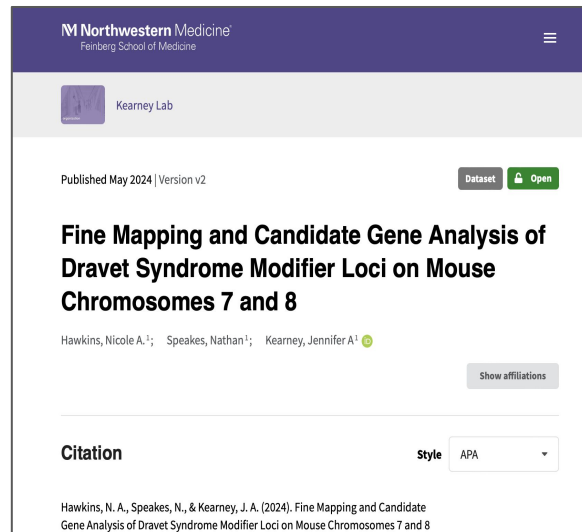
Dataset posted on 2024-02-07, 06:50 authored by [Hokan Kim](#), [William K. Epling](#), [Harry W. Abernathy](#), [Anthony Rollett](#), [Paul Salvador](#)

USAGE METRICS 


300	0	0
views	downloads	citations

The raw greyscale microstructures (*.tif), meshed sub-volume input files (*.inp), and simulation output files (*.csv) are hosted at the linked pages. [Readme files](#) there explain the content.

 Is supplement to [Advantages of ionic conductors over electronic](#)




Northwestern Medicine
Feinberg School of Medicine

 Kearney Lab

Published May 2024 | Version v2 Dataset Open

Fine Mapping and Candidate Gene Analysis of Dravet Syndrome Modifier Loci on Mouse Chromosomes 7 and 8

Hawkins, Nicole A.¹; Speakes, Nathan¹; Kearney, Jennifer A¹ 

Show affiliations

Citation Style APA ▾

Hawkins, N. A., Speakes, N., & Kearney, J. A. (2024). Fine Mapping and Candidate Gene Analysis of Dravet Syndrome Modifier Loci on Mouse Chromosomes 7 and 8



Implications for the Larger Data Sharing Landscape

Kristi Holmes, Northwestern University & Zenodo



Implications for the Larger Data Sharing Landscape

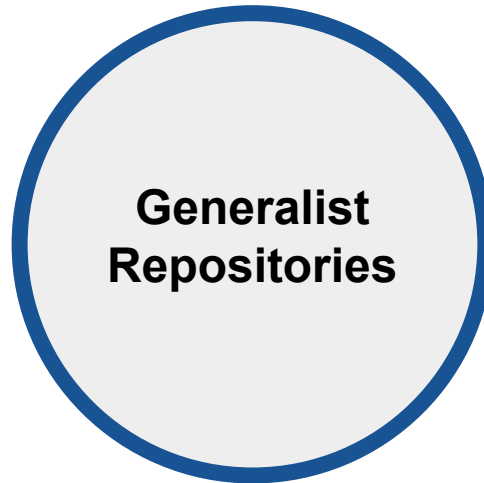
As the NIH data repository landscape grows with the adoption of the new NIH Data Management and Sharing Policy, there are **benefits and opportunities to global repositories working together** in this way to **meet the needs of research communities, funders, and institutions.**

In the future, discipline-specific data repositories and other research infrastructure providers may also wish to **adopt some of the common GREI capabilities to reduce the barriers to data sharing and support greater interoperability** across the repository landscape.



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.



Implications for the Larger Data Sharing Landscape

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.
- GREI repositories developed a recommendation of **common metadata fields** and **agreed to implement**
- Over half of the GREI repositories have **implemented Make Data Count standards** to date
- 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

“NIH believes the collective efforts of the GREI repository teams will help bring about the culture change required to share research data.”

~

Ishwar Chandramouliswaran

*Program Director, FAIR Data and Repositories
Office of Data Science Strategy (ODSS)
National Institutes of Health's (NIH)*



Support for a **culture** of data management and sharing that treats data sharing as a **standard scientific practice** that is valued and rewarded and furthers data reuse.





FORCE11

Audience Questions & Discussion

More information: Connect with GREI

- Join the **GREI Google Group** to receive updates on GREI outputs and events
<https://groups.google.com/u/1/g/contactgrei>
- Read the latest posts on the **GREI Community blog**
<https://medium.com/@blog-grei>
- Browse the resources and published outputs in the **GREI Community** on Zenodo
<https://zenodo.org/communities/grei/>
- Share your feedback To share questions or feedback, please get in touch via the **GREI Community Feedback Form** <https://forms.gle/n8uYdh4nB1LaLt497>

Find these slides at: doi.org/10.5281/zenodo.13152112

