



GREI

Generalist Repository Ecosystem Initiative

Data Sharing in Generalist Repositories

A Workshop by the NIH GREI repositories

September 19, 2023 1-3pm ET

NIH Research Festival 2023

Meet your workshop hosts



Ana Van Gulick, PhD

Government and
Funder Lead,
Figshare



Traci Snowden

Product Manager,
Elsevier



Sonia Barbosa

Manager of
Curation,
Harvard
Dataverse



Workshop Outline

- About GREI
- Introduction to Generalist Repositories in the NIH Data Landscape
- Best Practices for Sharing Data
- Best Practices for Finding Data
- Common Metadata Elements Across Repositories
- Examples of NIH Intramural Research Shared in Generalist Repositories
- Resources
- Feedback and Questions



Introduction to the Generalist Repository Ecosystem Initiative (GREI)



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.

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

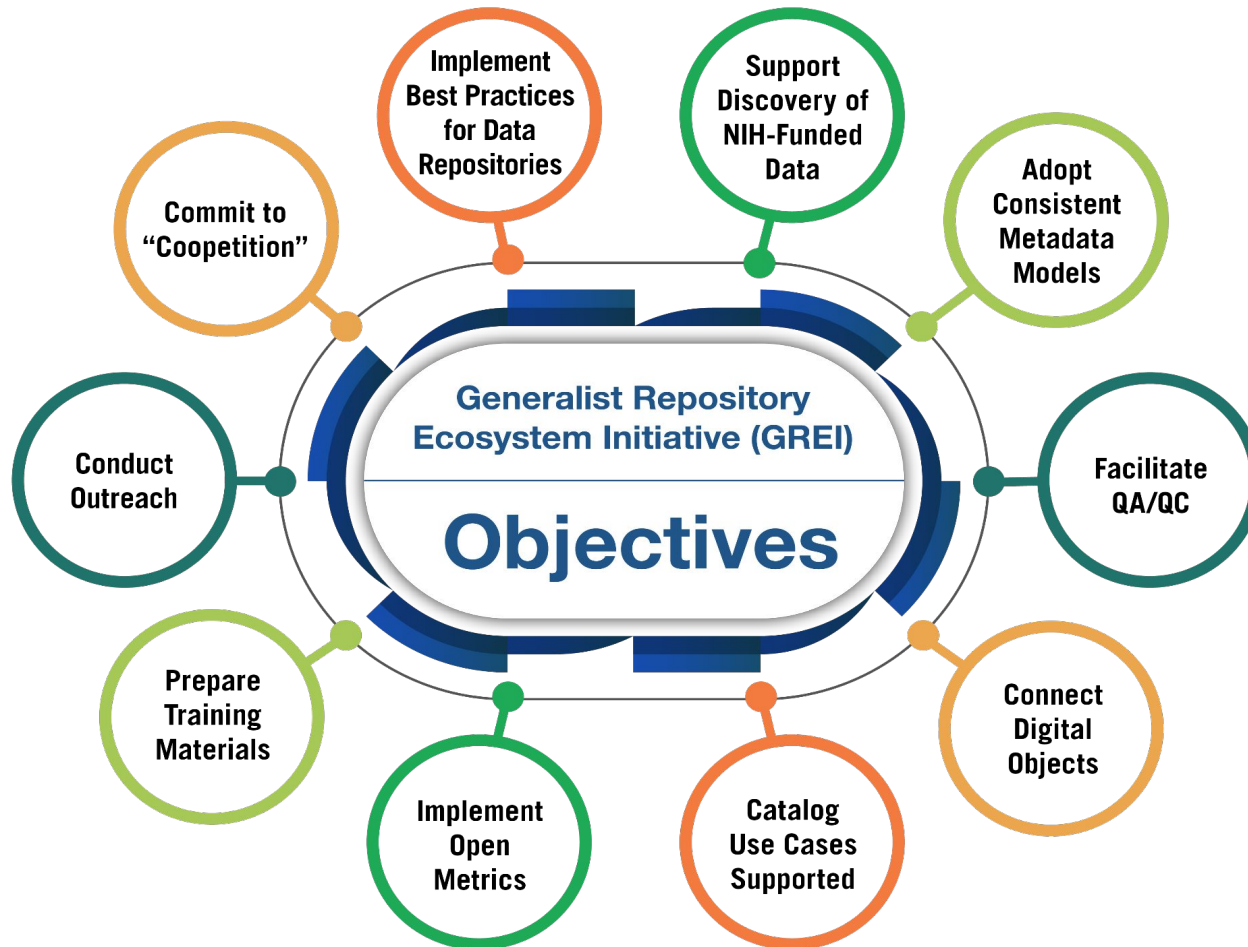
Encourage NIH-funded researchers to be both contributors and consumers to **increase the reuse of data**





Generalist Repository Ecosystem Initiative (GREI)





Discussion Question

~

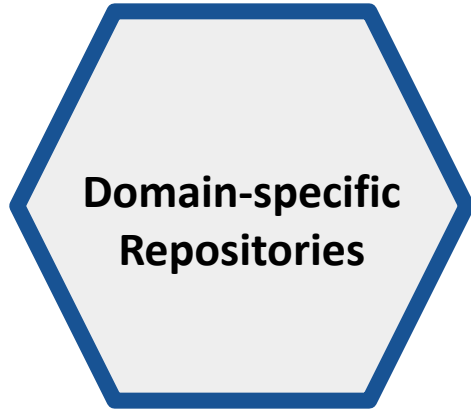
Getting to know you



Introduction to Generalist Data Repositories



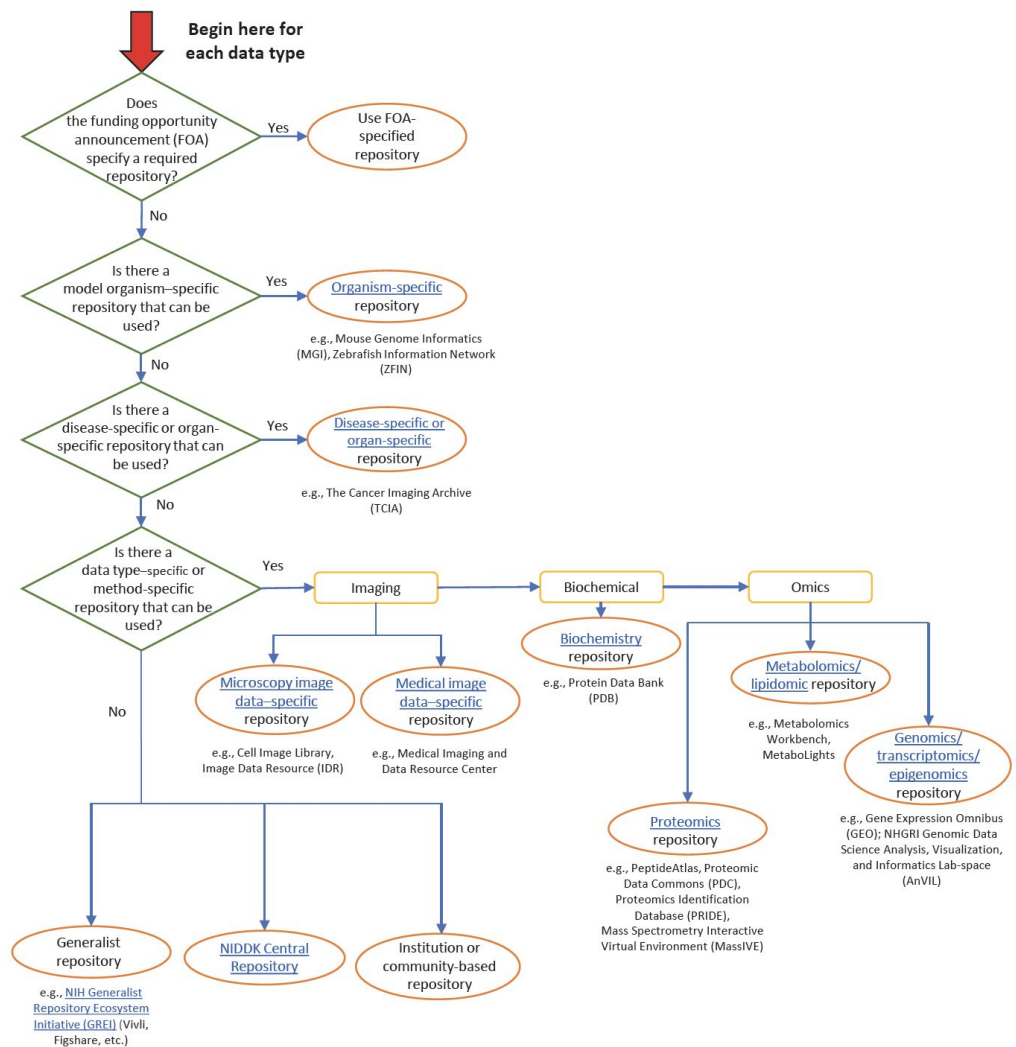
NIH Research Data Ecosystem



Considerations When Selecting a Repository

[Repository Selection Considerations Tool \(NIDDK\)](#) is intended to assist investigators to align the data types to be generated with appropriate repositories for submission and sharing.

After repository selection, investigators should confirm repository eligibility criteria and data requirements (e.g., data access type, preservation duration, storage capacity, possible data linkage strategies, acceptable file format)



NIH Desirable Characteristics of Data Repositories

When choosing a repository to manage and share data resulting from Federally funded research, here are some desirable characteristics to look for:

- **Unique Persistent Identifiers**
- **Long-Term Sustainability**
- **Metadata**
- **Curation and Quality Assurance**
- **Free and Easy Access**
- **Broad and Measured Reuse**
- **Clear User Guidance**
- **Security and Integrity**
- **Confidentiality**
- **Common Format**
- **Provenance**
- **Retention Policy**

[Guidance set forth by NIH](#)

And by [The National Science and Technology Council](#), cited in OSTP guidance



Repositories for Sharing Scientific Data

In general, NIH does not endorse or require sharing data in any particular repository, although some initiatives and funding opportunities will have individual requirements. **Overall, NIH encourages researchers to select the repository that is most appropriate for their data type and discipline. See [Selecting a Data Repository](#).**

Browse through this listing of NIH-supported repositories to learn more about some places to share scientific data. Note that this list is not exhaustive. Select the link provided in the "Data Submission Policy" column to find data submission instructions for each repository.

Can't find a repository that suits your data? Here are several more resources:

- A listing of [generalist repositories](#) that accepts all data types
- [Nature's Data Repository Guidance](#)
- [The Registry of Research Data Repositories](#)



See [Accessing Scientific Data](#) for more information about accessing data from NIH-supported repositories.

NIH-supported Scientific Data Repositories*



Institute or Center	Repository Name	Repository Description	Open Data Submission	Data Submission Policy	Open Time Frame for Data Deposit
All		Keyword Filter			
Common Fund	Metabolomics Workbench (MetWB)	The Metabolomics Program's Data Repository and Coordinating Center (DRCC), housed at the San Diego Supercomputer Center (SDSC), University of California, San Diego, has developed the Metabolomics Workbench. MetWB will serve as a national and international repository for metabolomics data and metadata and will provide analysis tools and access to metabolite standards, protocols, tutorials, training, and more.	Yes	How to submit data to MetWB	Yes
Common Fund	Stimulating Peripheral Activity to Relieve Conditions Portal (SPARC)	The SPARC Portal provides interactive access to a growing collection of data, maps, and computational studies that focus on the role of the autonomic nervous system in controlling organ function. These resources are made available to the public with the intent of advancing bioelectronic medicine towards more precise treatment of diseases and conditions.	Yes	How to submit data to SPARC	Yes



[Home](#) > [Data Management and Sharing Policy](#) > [Sharing Scientific Data](#) >

Generalist Repositories

While NIH encourages the use of domain-specific repositories where discipline or the type of data they generate, a generalist repository can be a useful place to share data. NIH does not recommend a specific generalist repository with a disciplinary focus. NIH does not recommend a specific generalist repository with a disciplinary focus.

- [Dataverse](#) 
- [Dryad](#) 
- [Figshare](#) 
- [IEEE Dataport](#) 
- [Mendeley Data](#) 
- [Open Science Framework](#) 
- [Synapse](#) 
- [Vivli](#) 
- [Zenodo](#) 

While NIH encourages the use of domain-specific repositories where possible, such repositories are not available for all datasets. When investigators cannot locate a repository for their discipline or the type of data they generate, a generalist repository can be a useful place to share data.

Generalist repositories accept data regardless of data type, format, content, or disciplinary focus. NIH does not recommend a specific generalist repository and the list below, which is not exhaustive, is provided as a guide for locating generalist repositories.



Use Case:



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.

Is this repository a good fit for my data?

- Data from any discipline (and interdisciplinary data)
- Data that has an appropriate home in a disciplinary or specialist repository
- Data in any format (and multiple formats)
- Data that should be shared in conjunction with software/code required for analysis
- Data that should be shared in conjunction with data in a disciplinary repository
- Data ready and intended for broad sharing and reuse
- Data and metadata that can benefit from quality control (curation)
- Data that requires a long-term embargo or managed access
- Data containing personally identifiable information (PII) or other sensitive content



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 COO 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

Common features and unique features

Common:
 Core Metadata
 Persistent Identifiers (PIDs)
 Discoverable
 Flexibility
 Open access, FAIR
 Metrics

Unique:
 Output types
 Storage, size limits
 Licenses
 Review
 Controlled Access
 Visualization
 Costs

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



[Home](#) > [Other Sharing Policies](#) > [NIH Institute and Center Data Sharing Policies](#)

NIH Institute and Center Data Sharing Policies

Data sharing is a priority across NIH. To this end, many institutes, centers, and research programs have instituted specific data sharing policies in addition to the trans-NIH policies. These policies are listed in the table below. Note that individual funding opportunities may specify other requirements or expectations, so be sure to read all instructions carefully.



Institute or Center	Data Sharing Policy Name	Description of Data Sharing Policy	Repositories
HEAL	HEAL Public Access and Data Sharing	Through the NIH HEAL Initiative Public Access and Data Sharing Policy (the Policy), NIH seeks to create an infrastructure that addresses the need for researchers, clinicians, and patients to collaborate on sharing their collective data and knowledge about opioid misuse and pain to provide scientific solutions to the opioid crisis. Under the Policy, applicants for extramural research funding (grants, cooperative agreements, contracts, and other transactions; "Applicants") for NIH HEAL Initiative Research Projects are required to submit a Public Access and Data Sharing Plan that (1) describes their proposed process for making resulting Publications and, to the extent possible, the Underlying Primary Data immediately and broadly available to the public or (2), if applicable, provides a justification to NIH if such sharing is not possible. Underlying Primary Data should be made as widely and freely available as possible while safeguarding the privacy of participants and protecting confidential and proprietary data.	Various HEAL-Compliant repositories
NCI	Cancer Moonshot™ Public Access and Data Sharing Policy	The primary goal of NCI's Cancer Moonshot™ is to significantly accelerate cancer research discovery and meaningful implementation. The Cancer Moonshot Public Access and Data Sharing Policy addresses the recommendation of the Blue Ribbon Panel's Enhanced Data Sharing working group to the National Cancer Advisory Board that researchers, clinicians, and patients should collaborate in sharing their collective data and knowledge about cancer to accelerate progress towards improving cancer outcomes. Under this policy, applicants for Cancer Moonshot Research Projects are required to submit a "Public Access and Data Sharing Plan" that describes their proposed process for making, to the extent possible, resulting Publications and the Underlying Primary Data immediately and broadly available to the public. Investigators applying for Cancer Moonshot funds must provide a justification to NCI if such sharing is not possible.	Genomic Data Commons , dbGaP , TCIA
NCI	NCI Clinical Trial Access Policy	NCI believes that the full value of NCI-supported Interventional Clinical Trials can be realized only if the results of clinical trials are published as rapidly as possible. The Clinical Trial Access Policy aims at ensuring public availability of results from NCI-supported clinical trials from all NCI-funded research grants, cooperative agreements, and/or contracts that support covered interventional clinical trials. Review the NCI Clinical Trial Access Policy for expectations of the policy.	Various
NHGRI	ENCODE Consortia Data Release, Data Use, and Publication Policies	Requires resource producers to release primary data along with an initial interpretation, in the form of genome features, to the appropriate public databases as soon as the data is verified. Consortia members will also identify validation standards that will be applied in subsequent analyses of the data or with additional experimentation where appropriate. All data will be deposited to public databases, such as GenBank or the ENCODE/ncf/ENCODE Data Coordination Centers (DCCs) and these pre-publication data will be available for all to use.	ENCODE , GenBank or other public databases
NHGRI	Genomic Data Sharing Policy	Broad data sharing promotes maximum public benefit from federally funded genomics research. NHGRI supports the broadest appropriate genomic data sharing with timely data release through widely accessible data repositories. These repositories may be open access (unrestricted) or, if more appropriate, controlled access. NHGRI encourages sharing of all data types. However, at this time the NIH GDS Policy and NHGRI implementation plans apply particularly to single nucleotide polymorphism (SNP) array data, genome sequence data, transcriptomic data, epigenomic data, or other molecular data produced by array-based technologies or high-throughput sequencing technologies. Data pertinent to the interpretation of genomic data — such as associated phenotype data (e.g., clinical information relevant to the disease under study), exposure data, and descriptive information (e.g., protocols or methodologies used) — are expected to be shared.	Widely accessible data repositories, including: dbGaP , AnVIL , SRA , GEO , ClinVar , National Cancer Institute Genomic Data Commons , UniProt , FlyBase , Database at the European Bioinformatics
NHLBI	NHLBI Policy for Data Sharing from Clinical Trials and Epidemiological Studies	Encourages all applicants to include a plan to address data sharing or to state why data sharing is not possible. For studies that meet the following criteria, applicants are required to provide a data sharing plan, which will be reviewed and approved by the relevant NHLBI program official: a) research applications/proposals requesting \$500000 direct costs; b) research studies that have 500 or more participants c) ancillary studies based on NHLBI-funded parent studies d) applications/proposals submitted in response to FOAs that specify inclusion of data sharing plans; or e) other research studies deemed appropriate for data sharing by NHLBI program official investigators.	NHLBI data repository through NHLBI and other repositories
NHLBI	NHLBI Supplement to the NIH Policy for Data Management and Sharing	Defers to the NIH DMS Policy scope and definitions in general but explicitly indicates scientific data include human and non-human data. This policy also sets additional NHLBI-specific expectations for compliance timelines, NIH-supported data repositories, and ancillary studies to NHLBI-funded parent studies.	NHLBI-supported researchers are expected to share scientific data through existing NIH-supported data repositories, such as NHLBI BioData Catalysts@IBDC . Generally discourages but allows for the deposition of the same data in multiple repositories with justification in the DMS Plan.

<https://sharing.nih.gov/other-sharing-policies/nih-institute-and-center-data-sharing-policies>



Home > [Other Sharing Policies](#) > [NIH Institute and Center Data Sharing Policies](#)

NIH Institute and Center Data Sharing Policies

Data sharing is a priority across NIH. To this end, many institutes, centers, and research programs have instituted specific data sharing policies in addition to the trans-NIH policies. These policies are listed in the table below. Note that individual funding opportunities may specify other requirements or expectations, so be sure to read all instructions carefully.



Institute or Center	Data Sharing Policy Name	Description of Data Sharing Policy
HEAL	HEAL Public Access and Data Sharing	Through the NIH... collective data and other transactions... the extent possible... made as widely available as possible.
NCI	Cancer Moonshot™ Public Access and Data Sharing Policy	The primary goal... recommendation... knowledge about... that describes the... Moonshot funds must provide a justification to NCI if such sharing is not possible.
NCI	NCI Clinical Trial Access Policy	NCI believes that the full value of NCI-supported interventional Clinical Trials can be realized only if availability of results from NCI-supported clinical trials from all NCI-funded research grants, cooperative Access Policy for expectations of the policy.
NCI	NCI Clinical Trial Access Policy	Final Trial Results are expected to be reported in a publicly accessible manner within twelve (12) months terminated earlier. Accordingly, data from incomplete trials are also expected to be reported within... achieve its primary aim. To comply with the Policy, Final Trial Results may be reported in a publicly... scientific journal, submitting study reports to publicly accessible registries dedicated to the dissemination available (e.g., abbreviated reports submitted to specific journals via simplified posting procedures).
NHGRI	ENCODE Consortia Data Release, Data Use, and Publication Policies	Requires resource producers to release primary data along with an initial interpretation, in the form... identify validation standards that will be applied in subsequent analyses of the data or with additional ENCODE/modENCODE Data Coordination Centers (DCCs) and these pre-publication data will be available.
NHGRI	Genomic Data Sharing Policy	Broad data sharing promotes maximum public benefit from federally funded genomics research. NIH repositories. These repositories may be open access (unrestricted) or, if more appropriate, controlled access. NHGRI encourages sharing of all data types. However, at this time the NIH GDS Policy and NHGRI transcriptomic data, epigenomic data, or other molecular data produced by array-based technologies. Data pertinent to the interpretation of genomic data — such as associated phenotype data (e.g., clinical methodologies used) — are expected to be shared.
NHLBI	NHLBI Policy for Data Sharing from Clinical Trials and Epidemiological Studies	Encourages all applicants to include a plan to address data sharing or to state why data sharing is not... which will be reviewed and approved by the relevant NHLBI program official: a) research application studies based on NHLBI-funded parent studies d) applications/proposals submitted in response to NHLBI program official investigators.
NHLBI	NHLBI Supplement to the NIH Policy for Data Management and Sharing	Defers to the NIH DMS Policy scope and definitions in general but explicitly indicates scientific data... timelines, NIH-supported data repositories, and ancillary studies to NHLBI-funded parent studies.

NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information

The NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information sets the expectation that all NIH-funded awardees and investigators conducting clinical trials will register and report results of their trial in [Clinicaltrials.gov](#). Read the details of the policy below.

Home > [Genomic Data Sharing Policy](#) > [About Genomic Data Sharing](#) > [Genomic Data Sharing Policy Overview](#)

Genomic Data Sharing Policy Overview

NIH expects the broad and responsible sharing of human as well as non-human genomic data resulting from NIH-funded research because the timely sharing of research results can accelerate discoveries that improve our ability to diagnose, treat, and prevent disease.

Other data sharing policies may also apply to your research.
[Learn More](#)

To comply with the [NIH Genomic Data Sharing Policy](#), NIH expects that investigators and institutions:

- Develop and provide a plan for sharing genomic data as a part of the [Data Management and Sharing Plan](#)
- Provide an Institutional Certification form at Just-in-Time, if working with human data
- Submit genomic data in a timely manner to an appropriate repository
- Responsibly use controlled-access data
- Appropriately cite controlled-access data in publications and presentations

Individual NIH Institutes, Centers, or Offices may have additional expectations (see [GDS Policy Expectations by NIH Institute & Center](#)).

Select each step below to learn more.

[BEFORE SUBMITTING GENOMIC DATA](#)
[SUBMITTING GENOMIC DATA](#)
[ACCESSING GENOMIC DATA](#)


HEAL Public Access and Data Sharing

NIH
HEAL
INITIATIVE

Data
Stewardship
Group

[ABOUT](#)[CALENDAR](#)[RESOURCES](#)

HEAL Data Repository Selection Guide

HEAL investigators are, in most cases, required to [share their HEAL-funded data](#), but the HEAL Data Platform itself does not host data. Instead, the HEAL Data Ecosystem uses appropriate, established long-term data storage repositories. To help HEAL investigators select a repository and meet their data sharing and [FAIR data practices](#) obligations, the HEAL Data Stewards evaluated and selected a number of HEAL-compliant repositories. The Stewards chose these compliant repositories from the [NLM Data Sharing Resources](#) lists and evaluated them based on [HEAL data repository selection principles](#).

Generally, HEAL investigators who have access to a repository managed by their administering NIH Institute or Center (IC) should explore that option first. For further information, please read the NIH IC Section below. If this is not applicable, investigators should review the list to identify a repository well suited to the organism they are studying, the type of data they are producing, and/or a repository they have used in the past. Investigators should submit data to a discipline-specific repository, where possible, but may submit to a [generalist repository](#) if their data do not fit naturally into one of the compliant domain-specific repositories. If you feel the best repository for your data is not on the HEAL-compliant list, please [contact the HEAL Stewards](#).

With HEAL investigator input, the Stewards will continue to work together with NIH to evaluate and update the list based on HEAL's needs.



HEAL Data Repository Selection Guide

HEAL-Compliant Repository List

The repositories below represent the most up-to-date list of compliant destinations for HEAL data. For more information about each repository, we provide direct links to repositories and re3data/FAIRsharing entries. Re3data is a global registry of research data repositories, providing objective summaries of repository attributes. FAIRsharing is a similar resource, providing curated and crowd-sourced descriptions and evaluations of repository features.

The HEAL Stewards ordered this list loosely by preference. In many cases, you should go down the list and select the first repository to which you are eligible to submit data and that accepts your data type.

Some NIH Institutes or Centers (ICs) require their HEAL investigators, or a subset of their HEAL investigators, to deposit their data in a repository managed by the IC. Be sure to check your Notice of Award language to see if your administering IC has a supporting repository.

☰ COLUMNS ⚙ FILTERS ≡ DENSITY ⬇ EXPORT

Repository	Descriptive Tags	Organism	IC/Program	Get Started Here	Overview
Vivli	Clinical Trial, Generalist	Human	N/A	Vivli Guidance	View re3data entry
MassIVE	Mass spectrometry	Unrestricted	N/A	MassIVE Guidance	View re3data entry
National Sleep Research Resource (NSRR)	Sleep recordings and r...	Human	N/A	NSRR Guidance	View re3data entry
Figshare	Generalist	Unrestricted	N/A	Figshare Guidance	View re3data entry
Dryad	Generalist	Unrestricted	N/A	Dryad Guidance	View re3data entry
Dataverse	Generalist	Unrestricted	N/A	Dataverse Guidance	View re3data entry
Mendeley Data	Generalist	Unrestricted	N/A	Mendeley Data Workflow and Guidance	View re3data entry
OpenScience Framework	Generalist	Unrestricted	N/A	Open Science Framework Guidance	View re3data entry
Qualitative Data Repository at Syracuse University	Qualitative	Unrestricted	N/A	Qualitative Data Repository Guidance	View re3data entry

1-26 of 26 < >

Vivli
 Figshare
 Dryad
 Dataverse
 Mendeley Data
 Open Science Framework
 Zemodo



Research Resources

NIDDK Data
Management &
Sharing

- [NIDDK Guidance for Writing a DMS Plan](#)
- [NIDDK DMS Tools & Resources](#)

NIDDK Data Management & Sharing

The goal of the NIH Data Management and Sharing (DMS) policy is to maximize the availability of data from NIH-supported research to advance NIH's mission to enhance health, lengthen life, and reduce illness and disability. The DMS policy provides a consistent, minimum expectation of data management and sharing for all research supported by the agency. To help investigators implementing the NIH DMS policy, the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) offers Institute-specific guidance for drafting a DMS Plan, as well as tools and examples to supplement the [NIH DMS policy and resources](#) NIH.

Guidance for Writing a DMS Plan

Access NIDDK specific guidance for the DMS plan elements.

- [Data Type](#)
- [Related Tools, Software, and/or Code](#)
- [Standards](#)
- [Data Preservation, Access, and Associated Timelines](#)
- [Access, Distribution, or Reuse Considerations](#)
- [Oversight of Data Management and Sharing](#)

DMS Tools and Resources

Access NIDDK resources for completing a DMS Plan.

- [DMS Plan Worksheet](#)
- [DMS Plan Examples](#)
- [Data and Metadata Standards](#)
- [Selecting a Data Repository](#)
- [NIDDK DMS Webinar Series](#)
- [Frequently Asked Questions \(FAQ\)](#)
- [Glossary of DMS Terms](#)





Selecting a Data Repository





Using an appropriate data repository generally improves the FAIRness (Findability, Accessibility, Interoperability, and Reusability) of the data. Selection of an appropriate data repository is essential to maximize data sharing. NIDDK affirms the [desired repository characteristics](#) ^{NIH} established by NIH, and strongly encourages the use of existing repositories to the extent possible for preserving and sharing scientific data.

Investigators need to consider the type of data they will be submitting when selecting a repository. A short justification of the repository selected for each data type must be included.

NIDDK strongly encourages investigators to consider the factors below **in order** when selecting a repository:

1. NOFO requirement (e.g., NIDDK-funded, large, multi-site clinical studies should submit data to the NIDDK Central Repository).
2. Organism, domain, or data type-specific repositories.
3. Whether controlled access to data is required (e.g., for protection of human subjects' privacy).

Repository Selection Aides

- The [Repository Selection Considerations Tool \(NIDDK\)](#) (PDF, 293.45 KB)  is intended to assist investigators to align the data types to be generated with appropriate repositories for submission and sharing.
- [dkNET](#) ^{NIH} lists available data repositories used by NIDDK supported researchers and provides tools to help comply with data sharing requirements.
- [NIH Repositories for Sharing Scientific Data](#) ^{NIH}, [NIH-supported repositories and generalist repositories](#) ^{NIH} for a wide range of data types and disciplines.
- The NIDDK DMS webinar “[Finding a Repository for Your Data](#)” provided additional information about NIDDK-research relevant repositories. The webinar videos linked below provides:
 - an overview of the tools and resources available on the NIDDK Information Network (dkNET) through [Dr. Grethe's presentation](#) ^{NIH} .
 - data eligibility and acceptance criteria as summarized in a [presentation by Dr. Rodriguez](#) ^{NIH}  on the NIDDK Central Repository.
 - how generalist repositories can be leveraged when a domain or data-type specific repository is not available, as reviewed in the [presentation by Mr. Chandramouliswaran](#) ^{NIH} .





Selecting a Data Repository

Using an appropriate data repository generally imp (Accessibility, Interoperability, and Reusability) of t data repository is essential to maximize data sharir **repository characteristics** NIH established by NIH, existing repositories to the extent possible for pres Investigators need to consider the type of data they repository. A short justification of the repository se included.

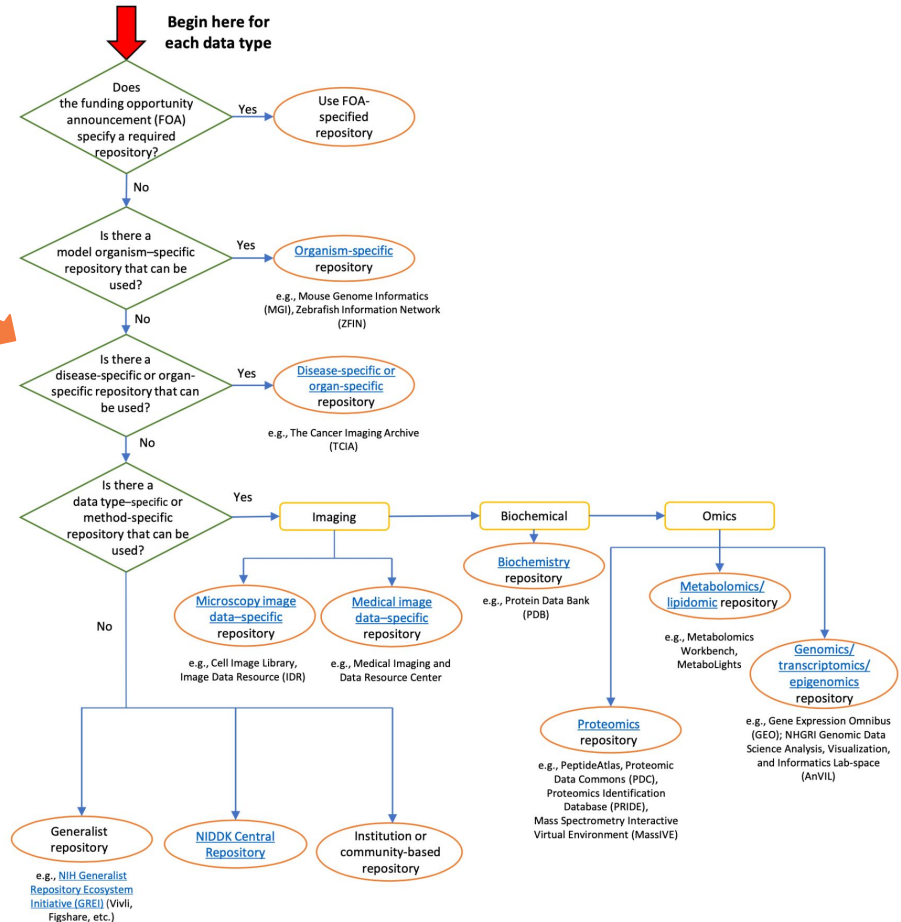
NIDDK strongly encourages investigators to consid selecting a repository:

1. NOFO requirement (e.g., NIDDK-funded, large, r submit data to the NIDDK Central Repository)
2. Organism, domain, or data type-specific reposit
3. Whether controlled access to data is required (e.g subjects' privacy).

Repository Selection Aides

- The [Repository Selection Considerations Tool](#) (N intended to assist investigators to align the data appropriate repositories for submission and sha
- [dkNET](#) lists available data repositories used by provides tools to help comply with data sharing r
- [NIH Repositories for Sharing Scientific Data](#) NIH [generalist repositories](#) NIH for a wide range of d
- The NIDDK DMS webinar “[Finding a Repository](#)” information about NIDDK–research relevant rep below provides:
 - an overview of the tools and resources avail Network (dkNET) through [Dr. Grethe’s pres](#)
 - data eligibility and acceptance criteria as su [Rodriguez](#) on the NIDDK Central Repo
 - how generalist repositories can be leverage specific repository is not available, as review [Mr. Chandramouliswaran](#)

Considerations When Selecting a Repository





Selecting a Data Repository

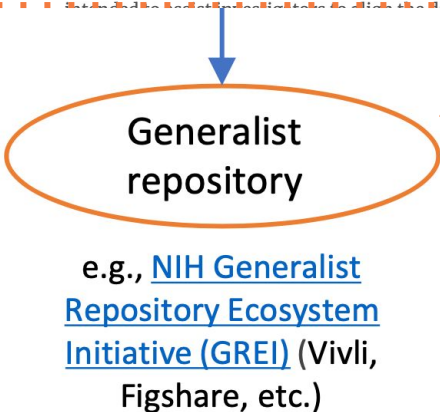
Using an appropriate data repository generally imp (Accessibility, Interoperability, and Reusability) of t data repository is essential to maximize data sharir **repository characteristics** NIH established by NIH, existing repositories to the extent possible for pres Investigators need to consider the type of data they repository. A short justification of the repository se included.

NIDDK strongly encourages investigators to consid selecting a repository:

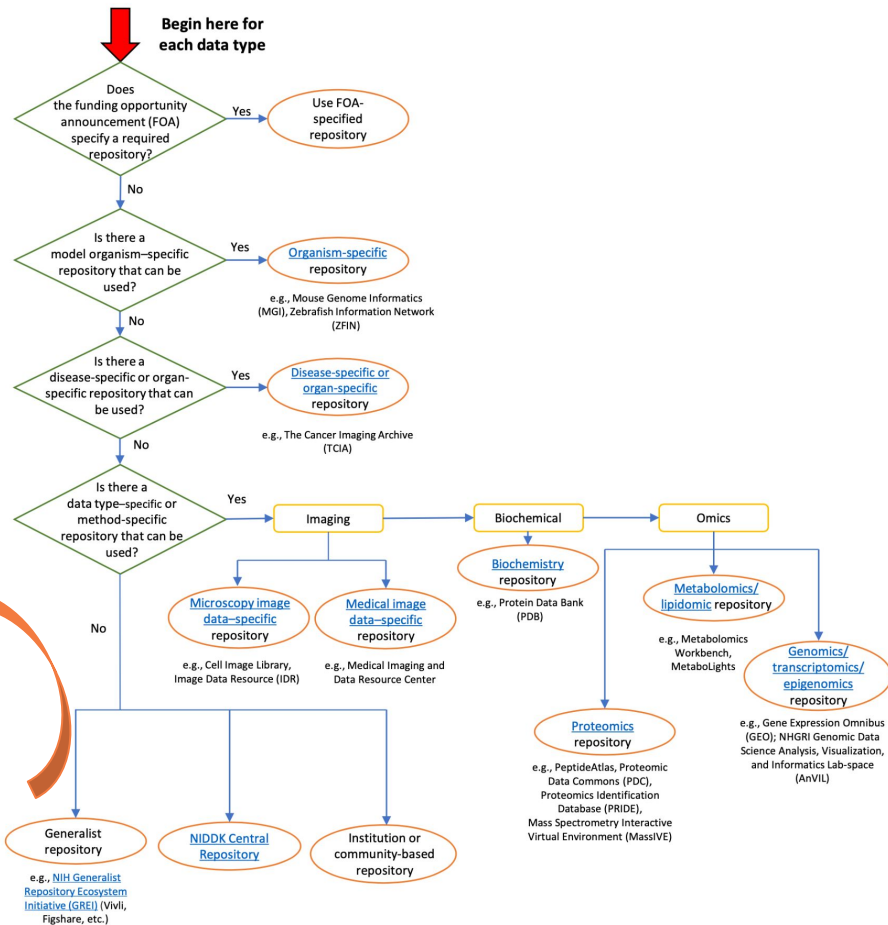
1. NOFO requirement (e.g., NIDDK-funded, large, r submit data to the NIDDK Central Repository).
2. Organism, domain, or data type-specific reposit
3. Whether controlled access to data is required (e.g subjects' privacy).

Repository Selection Aides

- The [Repository Selection Considerations Tool](#) (N



Considerations When Selecting a Repository



How is GREI making data sharing easier for intramural researchers?

- Flexible data sharing of many different output types
- Easy deposit workflows
- Combine with other repositories
- Interoperable common metadata and persistent identifiers
- Product enhancements for NIH data sharing and discovery
- Tool integrations

How is GREI improving discoverability of datasets across ALL repositories?



- Good discoverability across search engines and indexes
- Metadata for discovery and reuse
- Product enhancements for tracking and reuse of NIH data
- Common metrics
- Data Citations
- Credit for your work



Discussion Question

~

Experience using Generalist Repositories



Best Practices for Sharing Data

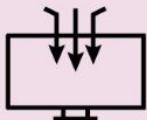


Best practices for sharing data

1

GATHER all stages of data needed for reanalysis

- Consider including the following:
 - Unprocessed raw data in recommended file types
 - Prepared and organized numerical data (tables, spreadsheets, etc.)
 - Code used to process and analyze data
 - Output (statistics and visualizations)



2

VERIFY files can be shared publicly

- Remove restricted materials such as:
 - Copyrighted or Licensed documents or software (CC0)
 - Content from published articles, grants, or patents
 - Data from third party with restricted terms-of-use
 - Identifiable human subjects data
 - Locations of endangered and vulnerable species



3

CHOOSE open file formats

- Use non-proprietary open file formats when possible to enable easy access, better preservation and interoperability.
- If you include proprietary files, consider also providing the data in an open format.
- Plain text formats are preferred.



4

ORGANIZE files logically

- Check files for errors or omissions.
- Name files descriptively and consistently.
- Omit needless files.
- Create a clear and logical file structure.
- Bundle organized files into compressed file archives.
- Try to keep individual files or archives smaller than 10GB.
- Verify file archives open and are not corrupted.



5

DESCRIBE your dataset in a README

- Write clearly for a broad audience.
- Describe processing pipeline and analysis steps.
- Define variables and allowable values.
- Describe software used to process, visualize, analyze, and compress your data (add open source recommendations if possible).



6

SHARE your data

- Identify a suitable data repository
- Review submission instructions
- Prepare your data following best practices
- Cite your published data package and share it on professional websites and/or social media



1



For ease of reuse and understanding include:

- Unprocessed, raw data
- Prepared & organized numerical data (spreadsheets, tables, etc.)
- Code, scripts, or software used to process and analyze data
- Output (statistics & visualizations)



2

Verify that files *can* be shared publicly



Remove restricted materials such as:

- Copyrighted or licensed documents or software (non-CC0)
- Content from published articles, grants, or patents
- Data from 3rd party with restricted terms-of-use
- Identifiable human subjects data
- Location information for species at risk (declining, threatened, endangered, or vulnerable status)



Choose **open** file formats



- Use non-proprietary, open file formats when possible to enable easy access, better preservation, and interoperability (commonly used in research community, unencrypted, uncompressed)

If you *do* include proprietary files, consider also providing the data in an open format
- Plain text formats are preferred (.csv)
- **License:** Ideally, CC0. This allows reuse without any restrictions, requirements



Organize files logically

A) Organized by File type

```
DatasetA.tar.gz
|- Data/
|  |- Processed/
|  |- Raw/
|- Results/
|  |- Figure1.tif
|  |- Figure2.tif
|  |- Models/
|- README.md
```

B) Organized by Analysis

```
DatasetB.tar.gz
|- Figure1/
|  |- Data/
|  |- Results
|     |- Figure1.tif
|- Figure2/
|  |- Data/
|  |- Results/
|     |- Figure2.tif
|- README.md
```

- Check files for errors or omissions
- Name files descriptively and consistently
- Remove unnecessary files
- Create a clear and logical file structure
- Bundle organized files into compressed file archives
- Keep individual files or archives less than 10GB in size
- Verify file archives open and are not corrupted



Prepare a **comprehensive** README file

- Write clearly for a broad audience
- Describe processing pipeline and analysis steps
- Define all variables, abbreviations, missing data codes, and units and allowable values
- Dates and locations of data collection (use standardized date formats)
- Provide description of file structure and contents
- Describe code, scripts, or software used to process, visualize, analyze, and compress the data

```
# Title of Dataset:
---

Brief summary of dataset contents, contextualized in experimental
procedures and results.

## Description of the Data and file structure

This is a freeform section for you to describe how the data are
structured and how a potential consumer might use them. Be as
descriptive as necessary. Keep in mind that users of your data might
be new to the field and unfamiliar with common terminology, metrics,
etc.

Describe relationship between data files, missing data codes, other
abbreviations used. Be as descriptive as possible.

## Sharing/access Information

Links to other publicly accessible locations of the data:

Was data derived from another source?
If yes, list source(s):
```



Decide where to share your data!

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

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 CC0 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

Repository Features

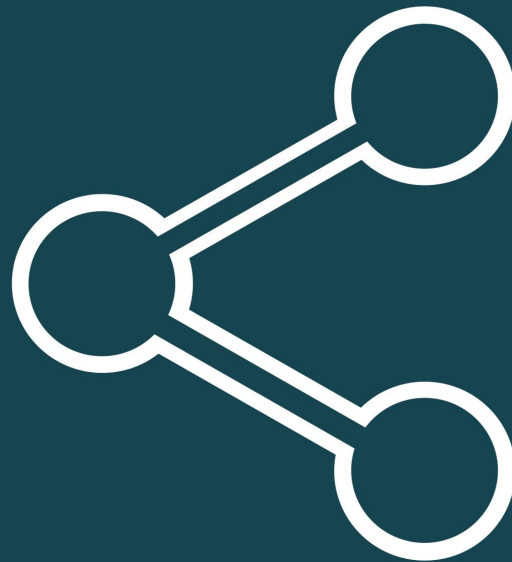
Common:
Core Metadata
Persistent Identifiers
Discoverable
Flexibility
Open access, FAIR
Metrics

Unique:
Output types
Storage, size limits
Licenses
Evaluation
Controlled Access
Visualization
Costs



In Review: Best practices for sharing data

- 1 Gather all data needed for reanalysis
- 2 Verify files can be shared publicly
- 3 Choose open file formats
- 4 Organize files logically
- 5 Describe your dataset in a detailed README file
- 6 Choose a suitable repository to share your data

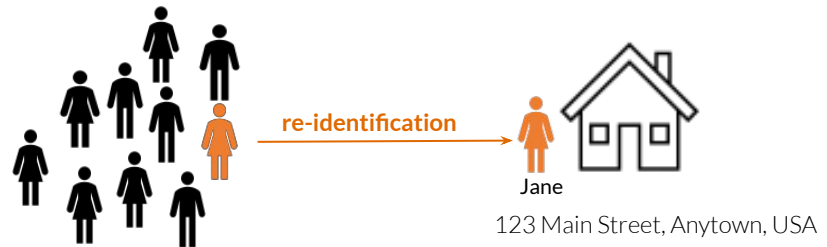


Best practices for sharing data in a **generalist repository**

Data preparation:

Additional considerations for **human subjects data**

- What level of data transformation does the repository require for reducing risk of re-identification?
- What privacy services or support does your institution offer?
- Governance: Do data requests go through a managed approval process or can anyone get at the data without restriction?
- Technical: Is the data accessible only in a secure enclave or workspace or can users download the data to their own computers?



Best practices for sharing data in a generalist repository

Example of de-identified human subjects data shared in Harvard Dataverse

Terms of use and application form for data requests

Children of Gay Fathers, 1985
Version 3.2

Bozett, Frederick W., 2022, "Children of Gay Fathers, 1985", <https://doi.org/10.7810/ONND24WQ>, Harvard Dataverse, V3

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

Description

The purpose of this research was to discover how children cope with having a gay father. Participants were 19 children of gay fathers, living in Iowa City, IA, Oklahoma City, OK, or San Francisco, CA. Of the children, 13 were female and 6 were male. The children's ages ranged from 14 to 35.

Data were collected through in-depth, unstructured interviews. Questions asked included how the children found out about their father's homosexuality, how they felt about it, whom they have told and why, whom they haven't told and why, what other people's reactions have been, whether their relationship with their father is any different since they found out that he is gay, whether their father acts any differently toward them now that his homosexuality is out in the open, and the advantages and disadvantages of having a gay father.

The Murray Research Archive holds copies of the 19 interview transcripts.

Subject
Social Sciences

License/Data Use Agreement
Custom Dataset Terms

Files | Metadata | Terms | Versions

Search this dataset...

Restricted, de-identified data

00874Bozett-Children-Measures.pdf
Adobe PDF - 1.0 MB
Published Aug 6, 2018
0 Downloads
MDS: e4d...c4a
Collection of blank measures used in the study
1. Documentation

00874Bozett-Children-MemoOfAgreement.pdf
Adobe PDF - 375.9 KB
Published Nov 27, 2007
30 Downloads
MDS: fc7...e44
Legal agreement between data depositor and Murray Archive
4. Detailed Usage Terms

00874Bozett-Children-PaperData-Subject01-Female.pdf
Adobe PDF - 718.9 KB
Published Aug 23, 2018
0 Downloads
MDS: fc...34c
Subject 01 Interview Transcripts
2. Data

00874Bozett-Children-PaperData-Subject02-Female.pdf
Adobe PDF - 1.1 MB
Published Aug 23, 2018
0 Downloads
MDS: 0f6...bdf
Subject 02 Interview Transcripts
2. Data

00874Bozett-Children-PaperData-Subject03-Male.pdf
Adobe PDF - 804.0 KB
Published Aug 23, 2018
0 Downloads
MDS: eeb...bae
Subject 03 Interview Transcripts
2. Data

Dataset Terms

License/Data Use Agreement

Our **Community Norms** as well as good scientific practices expect that proper credit is given via citation. Please use the data citation shown on the dataset page.

Custom Dataset Terms — the following Custom Dataset Terms have been defined for this dataset.

Terms of Use

- The Murray Archive (the Distributor) has granted me a revocable license to use this dataset solely for the purposes of conducting research, and the Distributor may terminate this license at any time and for any reason.
- I will use the dataset solely for statistical analysis and reporting of aggregated information, and not for investigation of specific individuals or organizations, except when identification is authorized in writing by the Distributor.
- I will produce no links among the Distributor's datasets or among the Distributor's data and other datasets that could identify individuals or organizations.
- I represent that neither I, nor anyone I know, has any prior knowledge of the possible identities of any study participants in any dataset that I am being licensed to use.
- I will not knowingly divulge any information that could be used to identify individual participants in the study, nor will I attempt to identify or contact any study participant, and I agree to use any precautions necessary to prevent such identification.
- I will make no use of the identity of any person or establishment discovered inadvertently. If I suspect that I might recognize or know a study participant, I will immediately inform the Distributor, and I will not use or retain a copy of data regarding that study participant. If these measures to resolve an identity disclosure are not sufficient, the Distributor may terminate my use of the dataset.
- I will not reproduce the dataset except as is necessary for my scholarly purposes. I will destroy the dataset upon the completion of my scholarly work with it.
- I will not share data from the dataset (in any form or by any means) with any third party, including other members of my research team, as I understand that all users of data must obtain the data directly from the Distributor.
- I will make appropriate acknowledgement of the contributor of the dataset as well as the Distributor in any manuscript or presentation (published or unpublished) using the citation standard documented here: <http://hmdata.org/citation>
- THE DISTRIBUTOR MAKES NO WARRANTIES, EXPRESS OR IMPLIED, BY OPERATION OF LAW OR OTHERWISE, REGARDING OR RELATING TO THE DATASET.

Special Permissions

Submission of the following **Application For The Use Of Data** is required to access the data from this study.

Documentation unrestricted for access, to give users information on the project



Best practices for sharing data in a **generalist repository**

Leveraging persistent identifiers

- A persistent identifier is a long-lasting reference to a resource.
- Aiming to solve for lost and broken links to important information through both technical and human failures.
- Enables access to a resource even if it is moved to different locations or owners.
- Creates opportunities for interoperability with many kinds of infrastructure and systems.
- Key element in making data FAIR (Findable, Accessible, Interoperable, Reproducible)!

Here's what you can do:

- **Get and use PIDs**
- **Register for an ORCID iD!** → <https://orcid.org>
- **Include relevant related PIDs in the metadata for your software, dataset, and paper PIDs, even if your repository or publisher says they're optional.**
- **Use repositories that provide PIDs**



Best practices for sharing data in a generalist repository

There are PIDs for people, places, and things

PIDs for **people** (researchers) include ISNIs and ORCID iDs



Example:

<https://orcid.org/0000-0002-5989-8244>

PIDs for **places** (research organizations) include ROR and Funder IDs



Example:

<https://ror.org/05d5mza29>

PIDs for **things** (research outputs/inputs like grants, papers, projects, etc.) include Crossref and DataCite DOIs, IGSNs, and more!



Example:

<https://doi.org/10.17605/osf.io/jzu37>



Best Practices for Finding Data



Best practices for FINDING data in a generalist repository

Describing data for discovery and reuse

All GREI repositories support the DataCite metadata schema and may support others

Repositories will typically have some fields with built-in controlled vocabularies and others that are free-text

- **Title:** Be descriptive, include data type, location, timeframe, and subject*
- **Abstract:** Describe the data and context of its collection, not the results of a study*
- **Author name(s)** and **affiliation(s)**
- **Research domain**
- **Funding information, award number**
- **Keyword(s):** Add synonyms, broader or more specific terms, use standardized vocabularies where possible
- **Related works:** Link to supplemental info, data management plans, preprint, related article, links to other publicly accessible locations of the data, etc.



Google search for finding data

Example: Harvard Dataverse Abstract

Harvard Dataverse > Brain Genomics Superstruct Project (GSP) Dataverse >

Brain Genomics Superstruct Project (GSP)

Version 1.0.1

Buckner, Randy L.; Roffman, Joshua L.; Smoller, Jordan W., 2014, "Brain Genomics Superstruct Project (GSP)", <https://doi.org/10.7910/DVN/25833>, Harvard Dataverse, V10

Access Dataset +

Contact Owner Share

Dataset Metrics 13,081 Downloads

Description Large scale imaging data sets are necessary to address complex questions regarding the relationship between brain and behavior. The Brain Genomics Superstruct Project Open Access Data Release exposes a carefully vetted collection of neuroimaging, behavior, cognitive, and personality data for over 1,500 human participants. Each neuroimaging data set includes one high-resolution Magnetic Resonance Imaging (MRI) acquisition and one or more resting-state functional MRI acquisitions. Each functional acquisition is accompanied by a fully-automated quality assessment and pre-computed brain morphometrics are also provided.

Subject Other

License/Data Use Agreement Custom Dataset Terms

Files Metadata Terms Versions

[doi:10.7910/DVN/25833](https://doi.org/10.7910/DVN/25833)

By **subject matter** and **title**,
a DOI may be required

Search aggregators can support cross-repository search to understand all sources for a single dataset

Google

doi:10.7910/DVN/25833

Images Videos News Maps Shopping Books Flights Finance

About 362 results (0.41 seconds)

Harvard Dataverse
<https://dataverse.harvard.edu/dataset/DVN>
Brain Genomics Superstruct Project (GSP) - Harvard Dataverse
Persistent Identifier, doi:10.7910/DVN/25833. Publication Date, 2014-05-22. Title, Brain Genomics Superstruct Project (GSP).

Nature Journal
<https://www.nature.com/scientificdata/data-descriptors>
Brain Genomics Superstruct Project initial data release with ...
by AJ Holmes · 2015 · Cited by 339 — The GSP release data provides a carefully vetted collection of neuroimaging, behavioral, cognitive, and personality data for 1,570 participants.

By **author** and **title**
yields ample citations

Randy Buckner Brain Genomics

Images Videos News Shopping Books Maps Flights Finance

About 84,100 results (0.33 seconds)

Scholarly articles for **Randy Buckner Brain Genomics**
Brain Genomics Superstruct Project initial data release ... - Holmes - Cited by 339

Harvard Brain Science Initiative
<https://brain.harvard.edu/people/andy-buckner/>
Randy Buckner - Harvard Brain Science Initiative
Randy Buckner, PhD, Professor of Psychology and Neuroscience, Harvard University. Human Brain Network Organization & Genomics. Weighing about three pounds, the ...

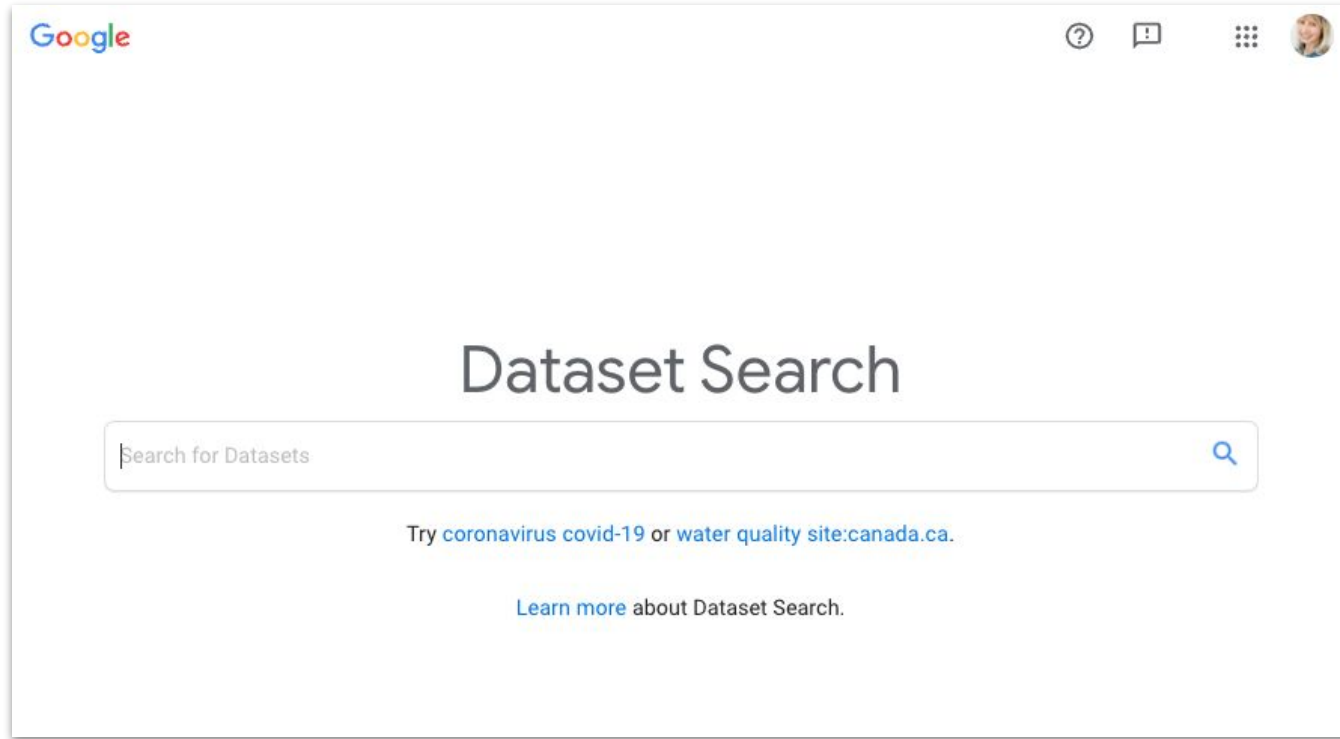
<https://psychology.fas.harvard.edu/people/andy-buckner/>
Randy L. Buckner | Department of Psychology
Randy Buckner received his BA in Psychology and his PhD in Neurosciences from Washington University in St. Louis. He is a member of the Center for Brain ...

Nature Journal
<https://www.nature.com/scientificdata/data-descriptors>
Brain Genomics Superstruct Project initial data release with ...
by AJ Holmes · 2015 · Cited by 339 — The present data fully cover the striatum, thalamus, and cerebellum allowing for analyses that extend beyond the cerebral cortex (see Buckner et ...

National Institutes of Health (.gov)
<https://pubmed.ncbi.nlm.nih.gov/>
Brain Genomics Superstruct Project initial data release with ...
by AJ Holmes · 2015 · Cited by 339 — Brain Genomics Superstruct Project initial data release with structural, functional, and behavioral measures ... Randy L Buckner. Affiliations. 1 Center f ...

Simons Foundation Autism Research Initiative
<https://www.sfi.org/funded-projects/the-brain-ge->
The Brain Genomics Superstruct Project

Leverage search & browse via **Google Dataset Search**!



Search functionalities in generalist repositories

- Keyword, Advanced & Faceted Search
- Search by Funding Agency, Content Type, Geographic Location, Journal, Institution
- Browse by Category, Featured Datasets, Collections
- Sort, Filter or Refine by Publication Year, Newest, Most Relevant, Field, File Extension, Funder

Add Search functionality for multiple repos; standard + key differentiators





Examples of search functionalities in generalist repositories



Search:

- Keyword search
- Advanced search
- Funding agency
- Faceted search
- Sorting
- Cross repository integration

Keyword Term

- AFRICA (328)
- AFRICA SOUTH OF SAHARA (309)
- EAST AFRICA (162)
- ASIA (148)
- health (145)

Funding Information Agency

- United States Agency for International Development (USAID) (247)
- Bill and Melinda Gates Foundation (BMGF) (77)
- Bill and Melinda Gates Foundation (31)
- World Bank (24)
- Bill & Melinda Gates Foundation (BMGF) (23)



Dataset details:

- Detailed dataset page
- Metadata
- Description
- Authors
- Citation
- Download options
- Documentation, README, code, etc.
- Multiple format download options

Browse:

- Categories/Subjects
- Featured Datasets/collections

ModulC_3rmember.tab
Data
Tabular Data - 202.7 KB
Published Aug 2, 2023
2 Downloads
14 Variables, 7045 Observations UNF6:q2N...11A

ModulA_hh.tab
Data
Tabular Data - 38.1 KB
Published Aug 2, 2023
0 Downloads
13 Variables, 1000 Observations UNF6:bnV...10A

Module1_5-5_var.tab
Data
Tabular Data - 416.2 KB
Published Aug 2, 2023
0 Downloads
21 Variables, 1732 Observations UNF6:MHOD...cgg

SurveyInstrument.pdf
Adobe PDF - 1.1 MB
Published Aug 2, 2023
0 Downloads
MDS: sc7...857

ModuleA_hh.tab
Data
Tabular Data - 38.8 KB
Published Aug 2, 2023
0 Downloads
13 Variables, 1200 Observations UNF6:bnV...10A

File Access

- Public
- Download Options
- Stata 14 Binary (Original File Format)
- Tab-Delimited
- RData
- Download Metadata
- Variable Metadata
- Data File Citation
- Export Options
- Data Explorer

IFPRI Dataverse Home
(International Food Policy Research Institute)

Harvard Dataverse >

Contact Share

The International Food Policy Research Institute (IFPRI) views the products of its research, including research datasets, as global public goods, and is committed to enabling their widespread distribution and use.

This is in keeping with the IFPRI Research Data Management and Open Access (RDMOA) Policy and the CGIAR Open Access and Data Management Policy.

The IFPRI Dataverse comprises datasets collected during the course of IFPRI research. These datasets include tables in various standard formats, survey instruments, codebooks, metadata, and other associated documentation. The Terms of Use require proper attribution of these datasets to IFPRI and any named authors.

Please direct questions about IFPRI datasets to IFPRI-Data or IFPRI-Library. For information on the latest resources and news on research data management, please visit IFPRI's Research Data website.

ASTI led by IFPRI
Agricultural Science and Technology Indicators (ASTI) Dataverse

Nexus SAMs led by IFPRI
Nexus SAM

HarvestChoice led by IFPRI
BETTER CHECKS, BETTER LIVES
IFPRI HarvestChoice Dataverse

IMPACT
International Model for Policy Analysis of Agricultural Commodities and Trade (IMPACT) Dataverse

Search this dataverse... Advanced Search

Datasets (8)
Files (13,127)

1 to 10 of 790 Results

Ethiopia Groundwater Games for Resource Governance **Embargoed**
Aug 2, 2023
International Food Policy Research Institute (IFPRI); Haramya University, 2023. "Ethiopia Groundwater Governance". <https://doi.org/10.21203/rs.3.rs-2723101/v1>. Harvest Dataverse, 31

Dataverse Category
Research Project (7)
Research Group (1)

Name (A-Z)
Name (Z-A)
Newest
Oldest

Ask the Data

File Tools Open in New Window

Ask the Data Want to know

View Data

What is this data about?

Answer please

This data is about households and their members

Browse: Recently added

File Tools Metadata Viewers

File Tools Open in New Window

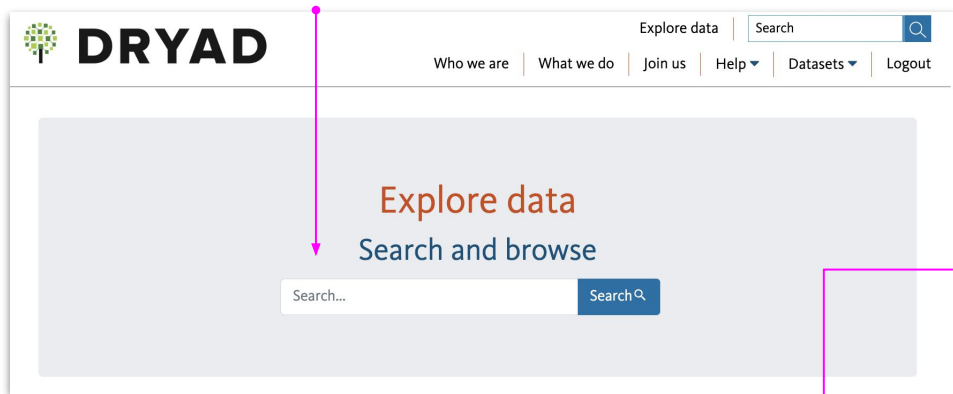
Tell me what you want to know
What is this data about?

Answer please

This data is about households and their members.

HHID	memberID	o01	o02	o03	o04	o05	o06	o07	o08	o09	o10	o11	o12
101001	1	2	1	23	1	2	4	1	1	0	1	1	1
101001	2	1	2	26	4	2	8	4	1	0	1	1	1
101002	1	2	1	25	4	2	4	3	1	0	1	1	1
101002	2	1	2	27	4	2	4	3	1	0	1	1	1
101002	3	1	3	5	1	1	1	1	0	1	1	1	1
101002	4	2	3	2	1	1	1	1	1	0	1	1	1
101003	1	2	0	76	1	2	5	3	1	2	2	1	1
101003	2	2	4	27	1	2	5	3	1	2	2	1	1
101003	3	1	3	6	1	1	1	1	1	2	2	1	1
101003	4	2	3	4	1	1	1	1	2	2	2	1	1

Free-text field to search & explore



Or, initiate a search by geographic location, subject keywords, journal, or institution

Find by...



Placename

North America, Europe, Australia, USA, Canada, United States, California, South America, Africa, [more »](#)



Subject keyword

Holocene, Anthropocene, Coevolution, Life History, Evolution, trade-offs, Inbreeding, Mate choice, present, Homo Sapiens, [more »](#)



Journal

Ecology and Evolution, PLOS ONE, Molecular Ecology, Evolution, Royal Society Open Science, Proceedings of the Royal Society B: Biological Sciences, Proceedings of the Royal Society B, Journal of Evolutionary Biology, The American Naturalist, [more »](#)



Institution

University of California, Davis, French National Centre for Scientific Research, University of California, Berkeley, University of Oxford, Cornell University, University of British Columbia, Chinese Academy of Sciences, University of Florida, University of Washington, [more »](#)



Search & browse

Additional options available to further refine search results by **file extension** and **funder**

Limit your search

- Subject keyword >
- Geographical Location >
- Journal >
- Institution ▾
 - University of California, Berkeley ✕ 776
 - University of California, Davis 48
 - University of California System 30
 - Cornell University 20
 - Stanford University 19
 - University of Florida 19
 - University of Minnesota 17
 - University of Washington 17
 - more »
- File Extension >
- Funder >

- 2015/16 El Niño increased water demand and pushed plants from a Mesic tropical montane grassland beyond their hydraulic safety limits >
- 2023 California Community Water System institutional type update >
- 2D attenuation model of Long Valley (CA) 🌐 >
- 3D attenuation model of Long Valley Caldera (CA) 🌐 >
- 3D microCT of lithium metal battery after charge and discharge >
- 5-species MHD study of Martian proton loss and source >
- A Bayesian method of evaluating discomfort due to glare: The effect of order bias from a large glare source >
- A fast machine-learning-guided primer design pipeline for selective whole genome amplification >
- A global phylogenomic study of the Thelypteridaceae >
- A highly contiguous genome assembly for the California quail (*Callipepla californica*) >





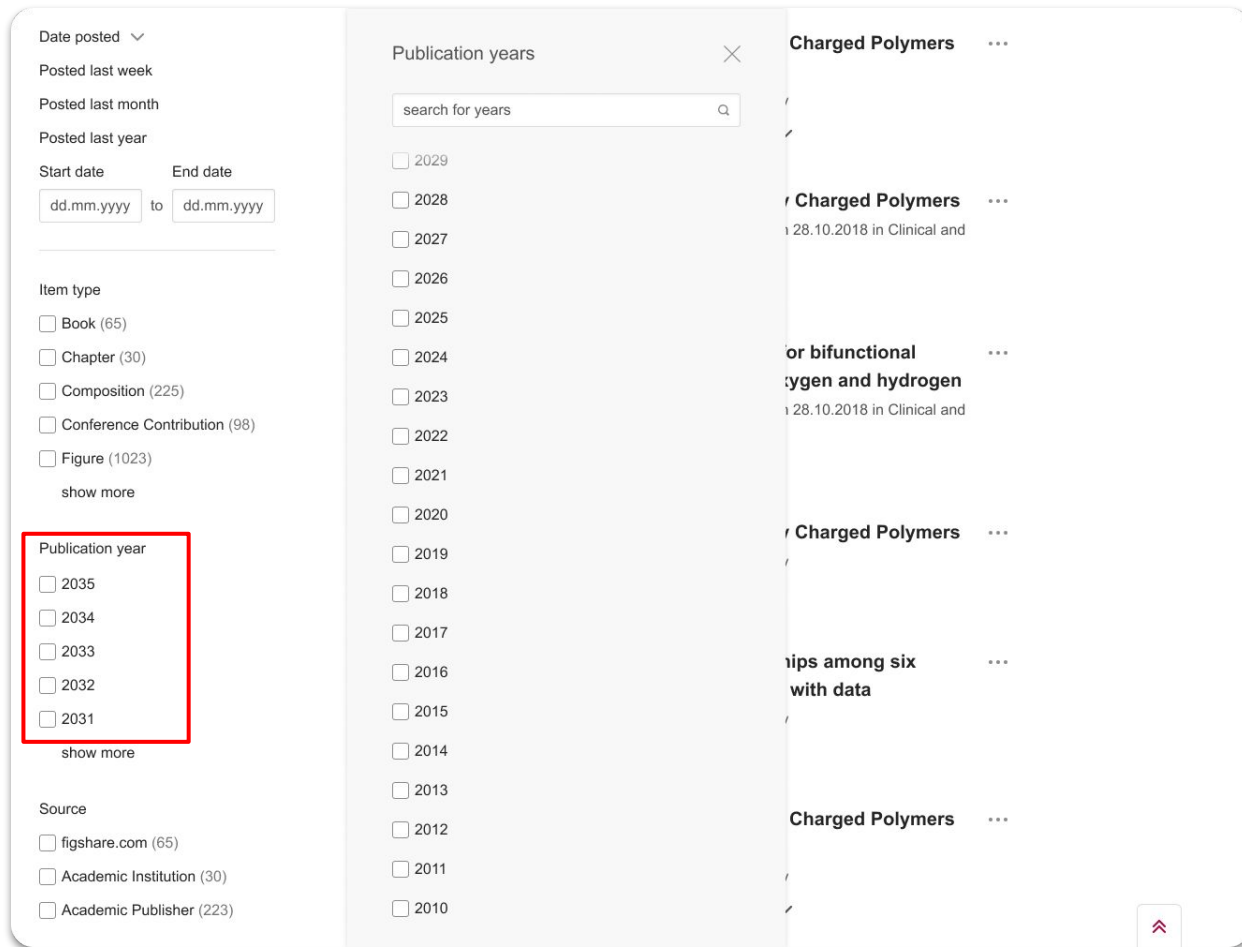
Improvements in planning:

- Provide more **robust search** options on Dryad site, include additional filters
- Offer search option **by date**, sort option
- Clean-up: remove ISSNs as titles, etc.
- Allow discoverability of **Research Areas** from metadata drop-down
- **Boolean** search functionality



Search & browse: publication year faceted search

Figshare is adding **publication year**
faceted search in August 2023.
<https://figshare.com/search/new>



The screenshot displays the Figshare search interface with several faceted search filters. The 'Publication year' filter is highlighted with a red box and contains the following options:

- 2035
- 2034
- 2033
- 2032
- 2031

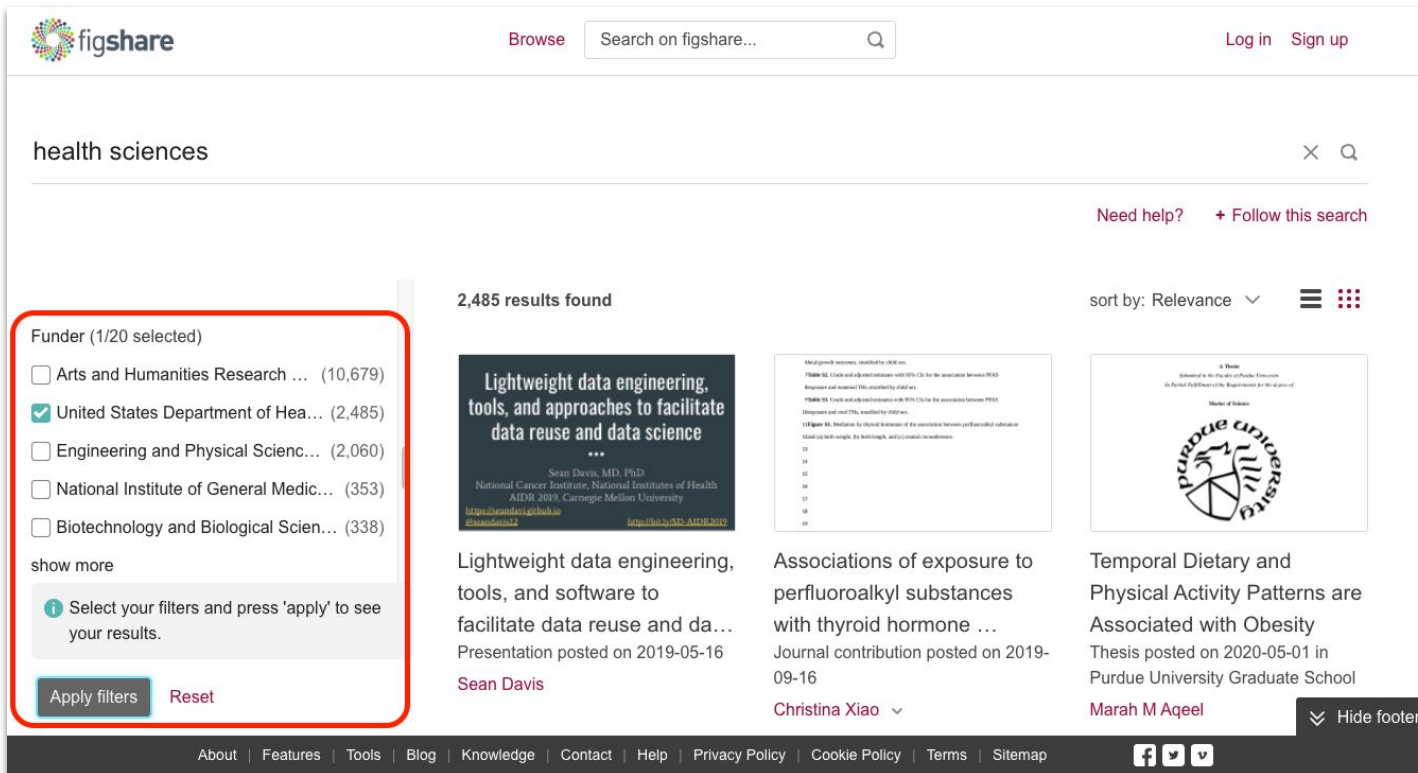
Other visible filters include:

- Date posted**: Posted last week, Posted last month, Posted last year.
- Start date** and **End date**: Input fields for dd.mm.yyyy to dd.mm.yyyy.
- Item type**:
 - Book (65)
 - Chapter (30)
 - Composition (225)
 - Conference Contribution (98)
 - Figure (1023)
- Source**:
 - figshare.com (65)
 - Academic Institution (30)
 - Academic Publisher (223)
- Publication years**: A list of years from 2010 to 2029 with checkboxes, and a search input field.

Search results on the right include entries such as 'Charged Polymers', 'Charged Polymers', 'or bifunctional hydrogen and hydrogen', and 'Charged Polymers'.



Search & browse: Filtering by funder



The screenshot shows the Figshare search results page for the query "health sciences". The page includes a search bar at the top with the text "Search on figshare...". Below the search bar, the results are filtered by "health sciences". A sidebar on the left shows a list of funders, with "United States Department of Health" selected. The main content area displays three search results, each with a thumbnail image and a title. The first result is "Lightweight data engineering, tools, and approaches to facilitate data reuse and data science" by Sean Davis. The second result is "Associations of exposure to perfluoroalkyl substances with thyroid hormone ..." by Christina Xiao. The third result is "Temporal Dietary and Physical Activity Patterns are Associated with Obesity" by Marah M Aqeel. The page also includes a "Need help?" link, a "Follow this search" button, and a "sort by: Relevance" dropdown menu.

health sciences

Need help? + Follow this search

2,485 results found

sort by: Relevance

Funder (1/20 selected)

- Arts and Humanities Research ... (10,679)
- United States Department of Health ... (2,485)
- Engineering and Physical Scienc... (2,060)
- National Institute of General Medic... (353)
- Biotechnology and Biological Scien... (338)

show more

Select your filters and press 'apply' to see your results.

Apply filters Reset

Lightweight data engineering, tools, and approaches to facilitate data reuse and data science

Sean Davis, MD, PhD
National Cancer Institute, National Institutes of Health
AIDR 2019, Carnegie Mellon University
<https://www.funder.org/hub.io> <http://dx.doi.org/10.1016/2019>

Lightweight data engineering, tools, and software to facilitate data reuse and da...
Presentation posted on 2019-05-16
Sean Davis

Associations of exposure to perfluoroalkyl substances with thyroid hormone ...
Journal contribution posted on 2019-09-16
Christina Xiao

Temporal Dietary and Physical Activity Patterns are Associated with Obesity
Thesis posted on 2020-05-01 in Purdue University Graduate School
Marah M Aqeel

Hide footer

About | Features | Tools | Blog | Knowledge | Contact | Help | Privacy Policy | Cookie Policy | Terms | Sitemap

Figshare added funder as a search facet in May 2023.



Users can filter by **Publication Year** using the slider feature or manual date entry



Filter Results

PUBLISHED DATE ^

1970 2023

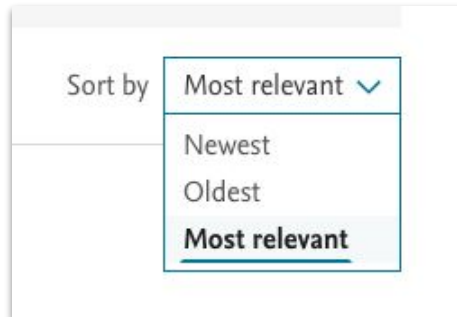
1970 2023

From To

1970 2023

The image shows a 'Filter Results' section with a 'PUBLISHED DATE' filter. It features a horizontal slider with blue circular handles at 1970 and 2023. Below the slider are two input fields labeled 'From' and 'To', both containing the values '1970' and '2023' respectively. An upward-pointing arrow is visible to the right of the filter title.

Publication date may also be sorted in order by newest to oldest, or based on relevance



Sort by

Most relevant v

Newest

Oldest

Most relevant

The image shows a 'Sort by' dropdown menu. The menu is open, displaying four options: 'Most relevant' (with a downward arrow), 'Newest', 'Oldest', and 'Most relevant' (which is highlighted with a blue background and a red underline).



Filter by **content type** by checking the desired check-box. Number in parentheses (#) indicates number of items identified for each type based upon the search criteria.

This feature was incorporated into Mendeley prior to the GREI project.

From To

DATA TYPES ^

- Dataset (60)
- Collection (18)
- Image (15)
- Tabular Data (11)
- Document (10)
- Text (8)
- Other (4)
- Software/Code (4)
- Interactive Resource (3)
- Slides (3)
- File Set (2)
- Sequencing Data (1)

ABSTRACT INTRODUCTION: Acquired middle ear cholesteatoma can be classified as pr
is still controversial whether there is an association between the type of cholesteatoma a
association between hearing loss and the type of acquired cholesteatoma, and the status
historical cohort study involving patients diagnosed with acquired cholesteatoma who w
bone gaps and the status of the ossicular chain were analyzed for both types of cholestea

Dataset

Export: [APA](#) | [BibTeX](#) | [DataCite](#) | [RIS](#)

Hearing loss assessment in primary and secondary acquired cholesteatoma,
Olsen, Julia Maria, Ribeiro, Fernando de Andrade Quintanilha, Yasui, Mariana Miekio Me
Published 1 January 2022 | figshare Academic Research System

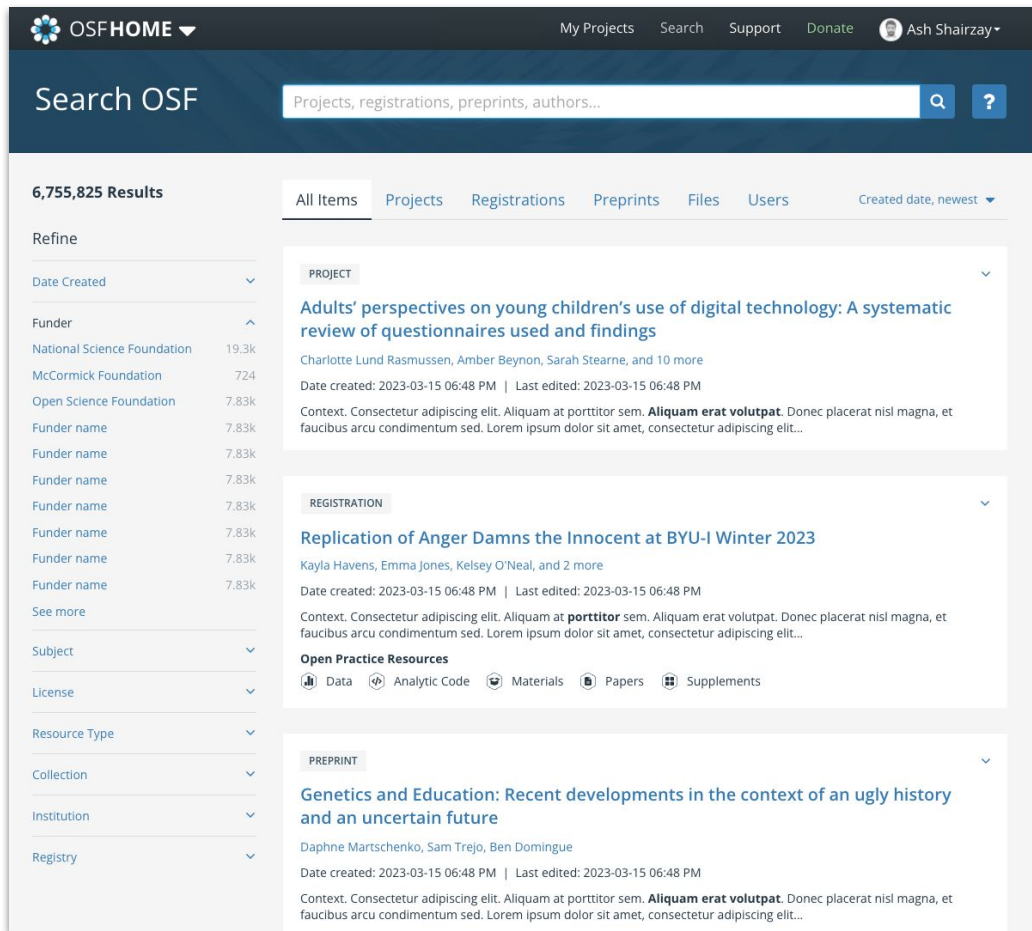
ABSTRACT INTRODUCTION: Acquired middle ear cholesteatoma can be classified as pr
is still controversial whether there is an association between the type of cholesteatoma a
association between hearing loss and the type of acquired cholesteatoma, and the status
historical cohort study involving patients diagnosed with acquired cholesteatoma who w
bone gaps and the status of the ossicular chain were analyzed for both types of cholestea

Dataset

Export: [APA](#) | [BibTeX](#) | [DataCite](#) | [RIS](#)



OSF Search & browse: Coming soon!



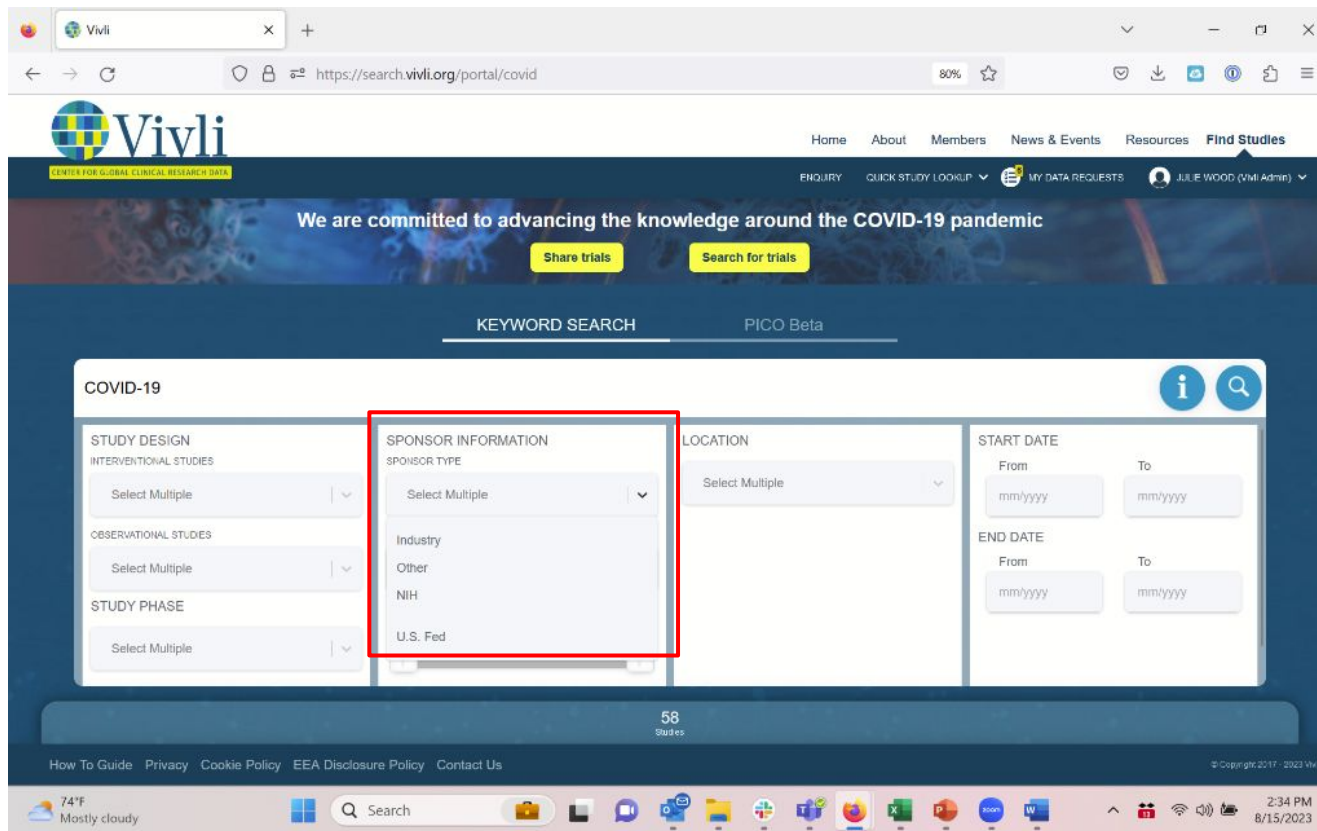
The screenshot displays the OSF search interface. At the top, the navigation bar includes the OSFHOME logo, a search bar with the placeholder text "Projects, registrations, preprints, authors...", and user information for "Ash Shairzay". Below the search bar, the results are categorized by type: "All Items", "Projects", "Registrations", "Preprints", "Files", and "Users". The "Projects" category is selected, showing a list of results. The first result is a project titled "Adults' perspectives on young children's use of digital technology: A systematic review of questionnaires used and findings" by Charlotte Lund Rasmussen, Amber Beynon, Sarah Stearne, and 10 more. The second result is a registration titled "Replication of Anger Damns the Innocent at BYU-I Winter 2023" by Kayla Havens, Emma Jones, Kelsey O'Neal, and 2 more. The third result is a preprint titled "Genetics and Education: Recent developments in the context of an ugly history and an uncertain future" by Daphne Martschenko, Sam Trejo, Ben Domingue. The left sidebar contains a "Refine" section with various filters such as "Date Created", "Funder", "Subject", "License", "Resource Type", "Collection", "Institution", and "Registry".

New search experience
releasing mid-August!

Learn more at:
<https://help.osf.io/>



Search & browse—ability to search via NIH



The screenshot shows the Vivli search portal for COVID-19. The browser address bar displays <https://search.vivli.org/portal/covid>. The page features a navigation menu with links for Home, About, Members, News & Events, Resources, and Find Studies. A user profile for Julie Wood (Vivli Admin) is visible. The main heading reads "We are committed to advancing the knowledge around the COVID-19 pandemic" with buttons for "Share trials" and "Search for trials". Below this, there are tabs for "KEYWORD SEARCH" and "PICO Beta". The search results are filtered for "COVID-19".

The search filters are organized into several sections:

- STUDY DESIGN:** Includes dropdowns for INTERVENTIONAL STUDIES, OBSERVATIONAL STUDIES, and STUDY PHASE, each with a "Select Multiple" button.
- SPONSOR INFORMATION:** This section is highlighted with a red box. It includes a "SPONSOR TYPE" dropdown menu with the following options:
 - Industry
 - Other
 - NIH
 - U.S. Fed
- LOCATION:** A dropdown menu with a "Select Multiple" button.
- START DATE:** Includes "From" and "To" date input fields (format: mm/yyyy).
- END DATE:** Includes "From" and "To" date input fields (format: mm/yyyy).

At the bottom of the search results area, it indicates "58 Studies". The footer contains links for "How To Guide", "Privacy", "Cookie Policy", "EEA Disclosure Policy", and "Contact Us". The copyright notice is "© Copyright 2017 - 2023 Vivli".

Search and browse with the sponsor type allowing for NIH studies (metadata provided by clinicaltrials.gov)





Phase IB/II Open-label Single Arm Study to Evaluate Safety and Efficacy of Tucatinib in Combination With Palbociclib and Letrozole in Subjects With Hormone Receptor Positive and HER2-positive Metastatic Breast Cancer

CLOSE

STUDY DESIGN
INTERVENTIONAL STUDIES

Select Multiple | v

OBSERVATIONAL STUDIES

Select Multiple | v

STUDY PHASE

Select Multiple | v

SPONSOR INFORMATION

SPONSOR TYPE

Select Multiple | v

SPONSOR

Select Multiple | v

SAMPLE SIZE

(Disabled)

LOCATION

Select Multiple | v

START DATE

From To

END DATE

From To

Phase IB/II Open-label Single Arm Study to Evaluate Safety and Efficacy of Tucatinib in Combination With Palbociclib and Letrozole in Subjects With Hormone Receptor Positive and HER2-positive Metastatic Breast Cancer

ID: NCT03054383 | 16-1661.cc
Condition or Disease: Breast Cancer
Intervention/treatment: Tucatinib in Combination with Palbociclib and Letrozole (Safety Cohort), Tucatinib in Combination with Palbociclib and Letrozole

Request Study v

View Study Details

Number enrolled: 42
Phase 1/Phase 2

The screenshot displays the Zenodo search interface. At the top, there is a search bar, navigation links for 'Upload' and 'Communities', and a user profile for 'kristiholmes@gmail.com'. The main content area shows search results for 'Found 8 results.' with a pagination control set to page 1. On the left, there are two filter panels: 'Access Right' with options for 'Open (429)', 'Restricted (17)', and 'Closed (5)'; and 'File Type' with various file formats like Pdf (298), Zip (69), Docx (13), Gz (13), Csv (12), Txt (11), Xlsx (9), Meta (7), Png (6), and Md (5). The search results list includes:

- December 20, 2022 (1.0.0) Dataset Open Access**: **Data and materials for "The Consequences of Data Dispersion in Genomics: A Comparative Analysis of Data Sources for Precision Medicine" manuscript** by García S. and Costa. Description: "Data and scripts for the "The Consequences of Data Dispersion in Genomics: A Comparative Analysis of Data Sources for Precision Medicine" manuscript" manuscript, sent to BMC Bioinformatics. Uploaded on December 20, 2022.
- October 28, 2022 (v1.0) Dataset Open Access**: **MIMESIS Paper Supplementary Dataset** by Romagnoli Dario and Benelli Matteo. Description: "This repository contains the code to perform all the analyses and the data generated by the author of the MIMeSis paper (Romagnoli et al., submitted). Uploaded on October 28, 2022.
- September 28, 2022 (1.0) Dataset Open Access**: **Interpretable AI for drug response prediction** by Li, Yihui, Hostallero, David, Emad, Amin. Description: "This dataset contains input data needed to run the models evaluated in our study "Interpretable deep learning architectures for improving drug response prediction: myth or reality?". The data required to run each model is organized into sub-folders, with the folder name being the correspon. Uploaded on September 28, 2022.
- April 6, 2021 (v1) Dataset Restricted Access**: **Dataset related to case report article "Gabapentin treatment in a patient with KCNQ2 developmental epileptic encephalopathy."** by Freri, Elena, Granata, Tiziana, Ragona, Francesca.

On the right side, there are two additional filter panels: 'Keywords' and 'Type'. The 'Keywords' panel shows filters for Precision Medicine (27), Bioinformatics (9), Machine Learning (9), Biodiversity (8), Taxonomy (8), Machine Learning (7), Personalized Medicine (6), Artificial Intelligence (5), Bioinformatics (5), and COVID-19 (5). The 'Type' panel shows filters for Publication (281) +, Dataset (69), Software (42), Presentation (19), Poster (17), Other (12), Image (7) +, and Video (4).

Refinement available across a wide range of fields <https://help.zenodo.org/guides/search/>, including all fields (e.g. funder, ORCID, datetime and ranges, and affiliation) identifier schemes, relations, and contributor types.



zenodo | Help

About Blog Help Projects Dev

Help FAQ Guides

Search guide

This guide explains how to perform advanced search queries on Zenodo using easy to understand examples.

Simple search (one or multiple terms)

Example: `open science`

Results will match records with the **terms** `open` or `science` in **any field**. Note that stemming is applied so e.g. `science` will also match `sciences`. Search results are ranked according to an algorithm that takes your query terms into account.

You can require **presence** of both terms using either the `+` or **AND** operator:

Examples: `+open +science` or `open AND science`

You can require **absence** of one or more terms using either the `-` or **NOT** operator:

Examples: `-open +science` or `NOT open AND science`

Phrase search

Example: `"open science"`

Results will match records with the **phrase** `open science` in **any field**.

Field search

Example: `title:open`

Results will match records with the **term** `open` in the **field** `title`. If you want to search for multiple terms in the title you must **group the terms** using parenthesis:

Example: `title:(open science)`

See the field reference below for the full list of fields you can search.

Combined simple, phrase or field search

Fields reference

The table below lists the data type of each field. Below is a quick description of what each data type means and what it is

- **string**: Field does not require exact match (example field: `title`).
- **string (exact)**: Field requires exact match (example field: `doi`).
- **CV**: Controlled vocabularies are similar to 'string (exact)' but takes only a specific list of values.
- **integer**: A number (range queries possible).
- **datetime**: An ISO8601 date or datetime, e.g. `2017-12-31` (range queries possible).
- **boolean**: `true` or `false`.

Field name	Type	Notes
<code>access_conditions</code>	string	
<code>access_right</code>	CV	See 'Access rights (CV)' below.
<code>alternate_identifier</code>	string (exact)	
<code>alternate_scheme</code>	CV	See 'Identifier schemes (CV)' below.
<code>communities</code>	string (exact)	Identifier of community.
<code>conceptdoi</code>	string (exact)	Related to DOI versioning.
<code>contributors.*</code>	string	Contributors name, affiliation, type and ORCID
<code>contributors.affiliation</code>	string	
<code>contributors.name</code>	string	
<code>contributors.orcid</code>	string (exact)	
<code>contributors.type</code>	CV	See 'Contributor types (CV)' below.
<code>created</code>	datetime	Creation timestamp of record in Zenodo.
<code>creators.*</code>	string	
<code>creators.affiliation</code>	string	
<code>creators.name</code>	string	
<code>creators.orcid</code>	string (exact)	
<code>description</code>	string	
<code>doi</code>	string (exact)	
<code>embargodate</code>	datetime	
<code>filecount</code>	integer	Number of files in record.
<code>filename</code>	string (exact)	
<code>filetype</code>	string (exact)	File extension (e.g. <code>pdf</code>).
<code>grants.*</code>	string	
<code>grants.acronym</code>	string	
<code>grants.code</code>	string (exact)	
<code>grants.funder.*</code>	string	
<code>grants.funder.acronym</code>	string	

Refinement available across a wide range of fields <https://help.zenodo.org/guides/search/>, including all fields (e.g. funder, ORCID, datetime and ranges, and affiliation) identifier schemes, relations, and contributor types.



Common Metadata Elements Across Repositories



Common metadata elements across repositories

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.

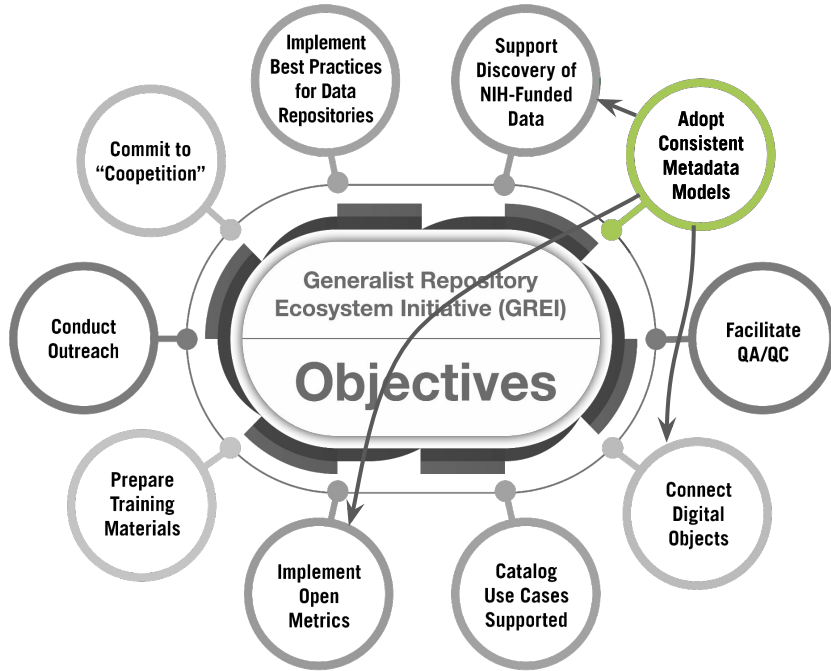
Curtin, Lisa; Feri, Lorenzo; Gautier, Julian; Gonzales, Sara; Gueguen, Gretchen; Scherer, David; Scherle, Ryan; Stathis, Kelly; Van Gulick, Ana, & Wood, Julie. (2023).

GREI Metadata and Search Subcommittee
Recommendations_V01_2023-06-29. Zenodo.
<https://doi.org/10.5281/zenodo.8101957>

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



Common metadata elements across repositories



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.



A persistent identifier (PID) is a unique, long-lasting reference to an entity.

<https://doi.org/10.5061/dryad.708gr>



[https://datadryad.org/stash/dataset/
doi:10.5061/dryad.708gr](https://datadryad.org/stash/dataset/doi:10.5061/dryad.708gr)

Special URL that is registered in a known system, like DOI, ORCID or ROR

Always points to the same resource (or a metadata representation)



Example: Dataset DOI - Harvard Dataverse

<https://doi.org/10.7910/DVN/DEAZAQ>



<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/DEAZAQ>



The screenshot shows the Harvard Dataverse interface for a dataset. At the top, the Harvard Dataverse logo and navigation links (Add Data, Search, About, User Guide, Support, Sign Up, Log In) are visible. The dataset title is "Replication Data for: The profile of research on Long-Covid: A survey". Below the title, there is a "Version 2.0" label and a thumbnail image of a network graph. The thumbnail includes the text: "Ren, Feng, 2022, 'Replication Data for: The profile of research on Long-Covid: A survey', https://doi.org/10.7910/DVN/DEAZAQ, Harvard Dataverse, V2" and a "Cite Dataset" link. To the right of the thumbnail are buttons for "Access Dataset" (with a dropdown arrow), "Contact Owner", and "Share". Below these are "Dataset Metrics" showing "0 Downloads". The "Description" section states: "The file contains raw data, data processing results, various data inspection results and a variety of estimation results. (2022-12-28)". The "Subject" is listed as "Social Sciences". The "License/Data Use Agreement" is "CC0 1.0". At the bottom, there are tabs for "Files", "Metadata", "Terms", and "Versions", and an "Export Metadata" button. A "Citation Metadata" section is expanded, showing the following information:

Field	Value
Persistent Identifier	doi:10.7910/DVN/DEAZAQ
Publication Date	2022-12-27
Title	Replication Data for: The profile of research on Long-Covid: A survey
Author	Ren, Feng (China University of Petroleum (East China))
Point of Contact	Use email button above to contact. Ren, Feng (China University of Petroleum (East China))
Description	The file contains raw data, data processing results, various data inspection results and a variety of estimation results. (2022-12-28)
Subject	Social Sciences
Depositor	Ren, Feng
Deposit Date	2022-12-27



How Recommendations Align to DataCite Metadata

We chose the DataCite metadata schema because:

- All GREI repositories already use it to register DOIs
- It's domain agnostic
- DataCite already collaborates closely with GREI
- Other services rely on metadata expressed in DataCite's schema, including metadata aggregators and DataCite's own Event Data service
- **The GREI Metadata Recommendations highlight** specific properties from the DataCite Metadata Schema (v4.4), beyond the minimum required fields.
- **Repositories are encouraged to** incorporate these properties in their metadata or identify a local equivalent field.
 - For example, an "Author Identifier" field may be mapped to the DataCite "nameIdentifier" sub-property of "Creator".
- **When registering a DOI with DataCite, recommended properties should** be included in the DataCite DOI metadata.



How Recommendations Align to DataCite Metadata



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.



So repositories need to collect information about who funded the research that produced the dataset

19	FundingReference	
19.1	funderName	
19.2	funderIdentifier	Use IDs from the CrossRef's Funder Registry or from ROR.
19.2.a	funderIdentifierType	Select "Crossref Funder ID" or "ROR" from DataCite's controlled list.
19.2.b	schemeURI	Use https://www.crossref.org/services/funder-registry/ or https://ror.org/
19.3	awardNumber	
19.3.a	awardURI	



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.



So repositories need to collect information about other research that cited and used the data

12	RelatedIdentifier	
12.a	relatedIdentifierType	Use controlled list values from DataCite.
12.b	relationType	Use controlled list values from DataCite.
12.f	resourceTypeGeneral	Use controlled list values from DataCite.



Example - search for datasets - DRYAD



Explore data

Search

Who we are | What we do | Join us | Help | Datasets | Logout

Dataset funder dashboard

Limit to

Partner institution:

Funder:

Submit

Reset

For dates

Date type:

Start date:

to

End date:

New dashboard available

With option to export results in CSV format

Title, Authors, DOI, Funder, Award, First submitted, Embargoed, Published

Title	Authors	DOI	Funder	Award	Submitted	Embargoed	Published
β-cell-specific deletion of Zfp148 improves nutrient-stimulated β-cell Ca2+ responses	Attie; Cardone; de Klerk; Emfinger; Foster; Gygi; Hebrok; Keller; Kendziorski; Kibbey; Lewandowski; Liu; Merrins; Mitok; Paulo; Perales; Rabaglia; Schueler; Simonett; Stapleton; Wang; Yu	10.5061/dryad.bcc2fqzcv	National Institute of Diabetes and Digestive and Kidney Diseases	R01DK101573	08/16/2021		05/11/2022
β-cell-specific deletion of Zfp148	Attie; Cardone; de Klerk; Emfinger; Foster; Gygi; Hebrok;	10.5061/dryad.bcc2fqzcv	National Institute of				

	A	B	C	D	E	F	G	H	I
1	Title	Authors	DOI	Funder	Funder id	Award	First submitted	Embargoed	Published
2	Data from: Age-at-injury	Green; Murphy; Ortiz; Ro	10.5061/dryad.5tbzrbp4r	National Institute of Neu	http://dx.doi.org/10.1303:R21NS120022		08/25/2021 06:01:03 UTC		01/01/2022 00:00:00 UTC
3	Data from: Use of an exc	Bowers; Carper; Muchoni	10.5061/dryad.7pvmcvdx	National Institute of Gen	http://dx.doi.org/10.1303:GM103440		12/26/2022 20:25:59 UTC		01/01/2023 00:00:00 UTC
4	Quantitative analysis of	Levit; Ryder	10.5061/dryad.h70rxwdg	National Institute of Gen	http://dx.doi.org/10.1303:SK12GM000680		07/06/2020 21:30:03 UTC		01/02/2021 00:00:00 UTC
5	Quantitative analysis of	Levit; Ryder	10.5061/dryad.h70rxwdg	National Institute of Gen	http://dx.doi.org/10.1303:1F32GM128407		07/06/2020 21:30:03 UTC		01/02/2021 00:00:00 UTC
6	Quantitative analysis of	Levit; Ryder	10.5061/dryad.h70rxwdg	National Heart, Lung, and	http://dx.doi.org/10.1303:5K22HL126922		07/06/2020 21:30:03 UTC		01/02/2021 00:00:00 UTC
7	Trans-specific polymorph	Lively; Million	10.5061/dryad.l76hdr82f	Eunice Kennedy Shriver N	http://dx.doi.org/10.1303:T32 H0049336		12/09/2021 01:41:04 UTC		01/02/2023 00:00:00 UTC
8	Data from: Total synthe	Adamson; Darzi; Donalds	10.5068/D1Q38N	National Institutes of Gen	http://dx.doi.org/10.13039:1000000057		10/19/2022 00:32:04 UTC		01/03/2023 00:00:00 UTC
9	Summary statistics from	Ambati; Faraco; Hallmay	10.5061/dryad.kd51c5b9l	National Institutes of He	http://dx.doi.org/10.1303:NIH-23724		10/29/2022 01:06:04 UTC		01/03/2023 00:00:00 UTC
10	Summary statistics from	Ambati; Faraco; Hallmay	10.5061/dryad.kd51c5b9l	National Institute of Mer	http://dx.doi.org/10.1303:5B2CMH08916		10/29/2022 01:06:04 UTC		01/03/2023 00:00:00 UTC

Example - search for datasets - Figshare

DataCite Commons

Pages ▾ Support

Sign In

Works People Organizations Repositories

4 Works

Accuracy of the InnovaveDX MTB/RIF test for detection of *Mycobacterium tuberculosis* and rifampicin resistance: a prospective multicentre study

Yunfeng Deng, Zichun Ma, Biyi Su, Guanghong Bai, Jianhua Pan, Quan Wang, Long Cai, Yanhua Song, Yuanyuan Shang, Pinyun Ma, Jing Li, Qianxuan Zhou, Gulibike Mulati, Dapeng Fan, Shanshan Li, Yaoju Tan & Yu Pang
 Dataset published 2023 in [figshare Academic Research System](#)

Early and accurate diagnosis of tuberculosis (TB) is necessary to initiate proper therapy for the benefit of the patients and to prevent disease transmission in the community. In this study, we developed the InnovaveDX MTB/RIF (InnovaveDX) to detect *Mycobacterium tuberculosis* (MTB) and rifampicin resistance simultaneously. A prospective multicentre study was conducted to evaluate the diagnostic performance of InnovaveDX for the detection MTB in sputum samples as compared with Xpert and culture. The calculated limit of detection (LOD) for InnovaveDX was 9.6 CFU/ml for TB detection and 374.9 CFU/ml for RIF susceptibility. None of the other bacteria tested produced signals that fulfilled the positive TB criteria, demonstrating a species-specificity of InnovaveDX. Then 951 individuals were enrolled at 7 hospitals, of which 607 were definite TB cases with positive culture and/or Xpert results, including 354 smear-positive and 253 smear-negative cases. InnovaveDX sensitivity was 92.7% versus bacteriologically TB standard. Further follow-up revealed that 61 (91.0%) out of 67 false-positive patients with no bacteriological evidence met the criteria of clinically diagnosed TB. Among 125 RIF-resistant TB patients diagnosed by Xpert, 108 cases were correctly identified by InnovaveDX, yielding a sensitivity of 86.4%. Additionally, the proportion of very low bacterial load in the discordant susceptibility group was significantly higher than in the concordant susceptibility group ($P = 0.029$). To conclude, we have developed a novel molecular diagnostic with promising detection capabilities of TB and RIF susceptibility. In addition, the discordant RIF susceptibility results between InnovaveDX and Xpert are more frequently observed in samples with very low bacterial load.

DOI registered January 2, 2023 via DataCite.



Dataset Biological sciences Mathematics Health sciences

<https://doi.org/10.6084/m9.figshare.21804085>

Publication Year

 2023 4

Work Type

 Dataset 4

License

 CC-BY-4.0 4

Field of Science

 Biological sciences 4

 Health sciences 4

 Chemical sciences 2

 Mathematics 2

Registration Agency

 DataCite 4


Example - search for datasets - OSF

OSF HOME

My Projects Search Support Donate Gretchen Gueguen

Search OSF

Enter search term(s) here

31 results

Refine

Funder: Deutsche Forschungsgemeinschaft x

Date created

Funder

Subject

License

Resource type

- JournalArticle 6
- Dataset 5
- Collection 1
- ConferenceProceeding 1
- Software 1

Provider

All Projects Registrations Preprints Files Users Relevancy

PROJECT

Development of a dynamic loudness model including perceptual weights (Entwicklung eines dynamischen Lautheitsmodells mit perceptiven Gewichten) - (DFG grants OB 346/6-1 and VE 373/2-1)

Alexander Fischenich , Jesko Verhey , Daniel Oberfeld , and Jan Hots

Date created: 2022-02-08 | Date modified: 2023-07-04

PROJECT

Does teachers' spirituality make a difference? The explanatory value of spirituality and religiosity for teachers' professional beliefs on teaching and learning

Alexander Christ , Nastja Häusler , Stephan Kröner , and Marcus Penthin

Date created: 2019-11-26 | Date modified: 2023-06-11

PROJECT

Characterization of the planarian surface electroencephalogram

Jannes Freiberg , Julian Keil , and Christian Kaernbach

Date created: 2023-04-15 | Date modified: 2023-04-21

BRAND NEW SEARCH!!!



Discussion Question

~

**What types of data are you sharing?
What are your data repository needs?**



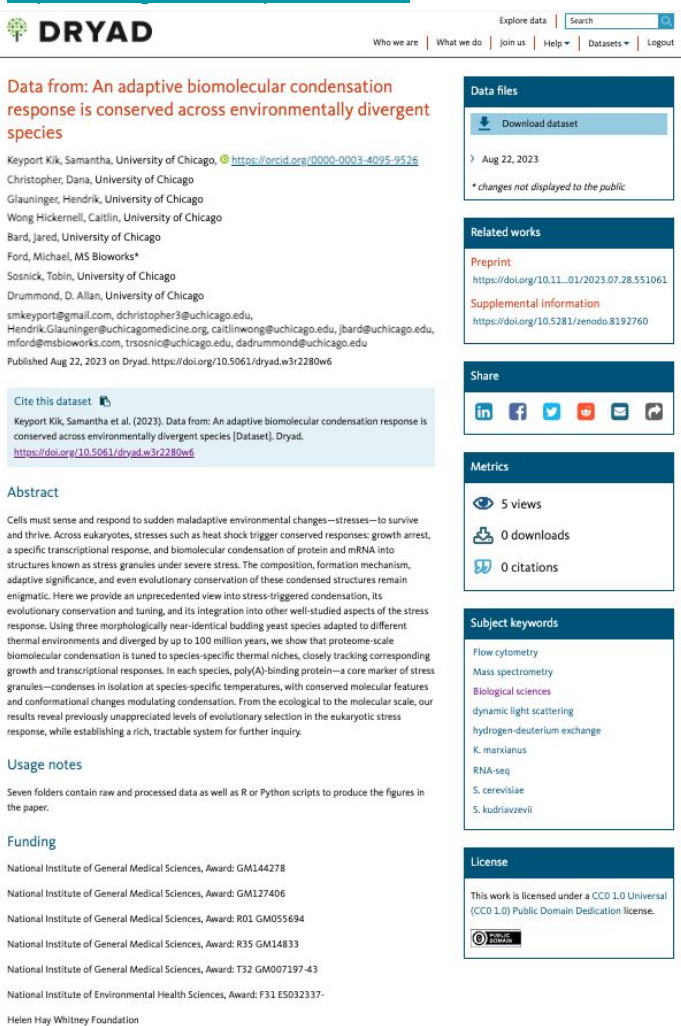
Examples of NIH Intramural Research Shared in GREI Repositories



Repository features that meet NIH requirements


- 1 All authors can connect their ORCID iDs
- 2 DOI assigned upon submission, does not change
- 3 Optional usage notes to help ensure ease of access & understanding
- 4 Funder & award details, searchable on the site
- 5 Search, browse, & explore datasets
- 6 Data package available for download with option to upload new versions, README is required
- 7 Links to preprint and supplemental info hosted on Zenodo; can also link to Data Mgmt Plan, related research, etc.
- 8 Metrics! Track views, downloads, and citations
- 9 Subject keywords (aid in discoverability); minimum 3 keywords required
- 10 CCO license required for all data hosted by Dryad permits unrestricted reuse, alternate license can be selected for Zenodo software files, supplemental files are CCBY

<https://doi.org/10.5061/dryad.w3r2280w6>



DRYAD Explore data | Search | Who we are | What we do | Join us | Help | Datasets | Logout

Data from: An adaptive biomolecular condensation response is conserved across environmentally divergent species

Keyport Kik, Samantha, University of Chicago,  <https://orcid.org/0000-0003-4095-9526>
Christopher, Dana, University of Chicago
Glauninger, Hendrik, University of Chicago
Wong Hickernell, Caitlin, University of Chicago
Bard, Jared, University of Chicago
Ford, Michael, MS Bioworks*
Sosnick, Tobin, University of Chicago
Drummond, D. Allan, University of Chicago
smkeyport@gmail.com, dchristopher3@uchicago.edu,
Hendrik.Glauninger@chicagogomedicine.org, caitlinwong@uchicago.edu, jbard@uchicago.edu,
mford@msbioworks.com, tsosnic@uchicago.edu, dadrummond@uchicago.edu
Published Aug 22, 2023 on Dryad. <https://doi.org/10.5061/dryad.w3r2280w6>

Cite this dataset

Keyport Kik, Samantha et al. (2023). Data from: An adaptive biomolecular condensation response is conserved across environmentally divergent species [Dataset]. Dryad. <https://doi.org/10.5061/dryad.w3r2280w6>

Abstract

Cells must sense and respond to sudden maladaptive environmental changes—stresses—to survive and thrive. Across eukaryotes, stresses such as heat shock trigger conserved responses: growth arrest, a specific transcriptional response, and biomolecular condensation of protein and mRNA into structures known as stress granules under severe stress. The composition, formation mechanism, adaptive significance, and even evolutionary conservation of these condensed structures remain enigmatic. Here we provide an unprecedented view into stress-triggered condensation, its evolutionary conservation and tuning, and its integration into other well-studied aspects of the stress response. Using three morphologically near-identical budding yeast species adapted to different thermal environments and diverged by up to 100 million years, we show that proteome-scale biomolecular condensation is tuned to species-specific thermal niches, closely tracking corresponding growth and transcriptional responses. In each species, poly(A)-binding protein—a core marker of stress granules—condenses in isolation at species-specific temperatures, with conserved molecular features and conformational changes modulating condensation. From the ecological to the molecular scale, our results reveal previously unappreciated levels of evolutionary selection in the eukaryotic stress response, while establishing a rich, tractable system for further inquiry.

Usage notes

Seven folders contain raw and processed data as well as R or Python scripts to produce the figures in the paper.

Funding

National Institute of General Medical Sciences, Award: GM144278
National Institute of General Medical Sciences, Award: GM127406
National Institute of General Medical Sciences, Award: R01 GM055694
National Institute of General Medical Sciences, Award: R35 GM14833
National Institute of General Medical Sciences, Award: T32 GM007197-43
National Institute of Environmental Health Sciences, Award: F31 ES032337-
Helen Hay Whitney Foundation

Data files

Download dataset

Aug 22, 2023

* changes not displayed to the public

Related works

Preprint
<https://doi.org/10.1101/2023.07.28.551061>

Supplemental information
<https://doi.org/10.5281/zenodo.8192760>

Share

Metrics

5 views
0 downloads
0 citations

Subject keywords

Flow cytometry
Mass spectrometry
Biological sciences
dynamic light scattering
hydrogen-deuterium exchange
K. marxianus
RNA-seq
S. cerevisiae
S. kudriavzevii

License

This work is licensed under a [CCO 1.0 Universal \(CCO 1.0\) Public Domain Dedication license](https://creativecommons.org/licenses/by/4.0/).



Zenodo supports preservation and sharing of a wide range of digital assets, including datasets, images, software, video and audio files, and more directly from <https://zenodo.org/>.

The screenshot shows the Zenodo interface for a dataset. At the top, the Zenodo logo and navigation links are visible. The dataset title is "Data from: An adaptive biomolecular condensation response is conserved across environmentally divergent species". Below the title, the authors are listed: Keyport Kik, Samantha; Christopher, Dana; Glauninget, Hendrik; Wong Hickernell, Caitlin; Bard, Jared; Ford, Michael; Sosnick, Tobin; Drummond, D. Allan. The dataset has 3 views and 0 downloads. A large orange arrow points to the "Data files" section, which is partially visible on the right side of the image. Below the title, there is a paragraph of text describing the study, followed by funding information and a list of files. The files list includes folders like "Protein" and "Flow cytometry" with sub-files such as "D19g1-v4.svg", "D19g1-v5.svg", etc., with their respective sizes.

The screenshot shows the Dryad website interface for the same dataset. The Dryad logo is at the top left. The dataset title is "Data from: An adaptive biomolecular condensation response is conserved across environmentally divergent species". Below the title, the authors are listed: Keyport Kik, Samantha, University of Chicago; Christopher, Dana, University of Chicago; Glauninget, Hendrik, University of Chicago; Wong Hickernell, Caitlin, University of Chicago; Bard, Jared, University of Chicago; Ford, Michael, MS Biowork; Sosnick, Tobin, University of Chicago; Drummond, D. Allan, University of Chicago. The dataset has 3 views and 0 downloads. A large orange arrow points to the "Data files" section, which is partially visible on the right side of the image. Below the title, there is a paragraph of text describing the study, followed by funding information and a list of files. The files list includes folders like "Protein" and "Flow cytometry" with sub-files such as "D19g1-v4.svg", "D19g1-v5.svg", etc., with their respective sizes.

Data from: An adaptive biomolecular condensation response is conserved across environmentally divergent species. Zenodo. <https://doi.org/10.5281/zenodo.8192760>



Mendeley Data Case Study 1:

Providing Steps to Reproduce a Dataset



Whole genome sequence data of *Leptospira weilii* and *Leptospira kirschneri* isolated from human subjects Sri Lanka

Published: 13 June 2023 | Version 1 | DOI: 10.17632/3f8tvpw348.1

Contributors: Indika Rathnabahu, Dinesha Jayasundara, Janith Warnasekara, Micheal Matthias, Joseph M. Vinetz, Suneth Agampodi

Steps to reproduce

Cultures were isolated using EMJH media

DNA extraction was performed using the DNeasy Blood & Tissue Kit, following a Gram-negative bacteria protocol, with an RNase cleanup step included.

The quantity of extracted DNA was measured using a Qubit 4 fluorometer.

High-quality genomic DNA (gDNA) was utilized to construct multiplexed PacBio SMRTbell libraries using the SMRTbell Express Template Prep Kit.

Shearing of 1 g of genomic DNA was achieved using Covaris g-tubes, and DNA concentration was enhanced using AMPure PB beads.

The DNA underwent repair and ligation to a barcoded 8A adaptor, followed by adherence to size selection instructions for Blue Pippin TM 4 kb or more,

Whole genome sequencing was conducted using the PacBio Single Molecule Real-Time (SMRT) platform

Raw data were processing and genome assembly using Canu 2.1 and Circlator and then circularized.

Genome annotation was conducted using RAST (Rapid Annotation Using Subsystem Technology) and NCBI Prokaryotic Genome Annotation Pipeline.

<https://data.mendeley.com/datasets/3f8tvpw348/1>



Step 1: Create New Draft DataSet and Add Title

Draft autosaved 14 September 2023 14:10 You can close and resume the draft whenever you want.

Draft (Version 1)

* Indicated fields which must be completed before publishing

Title *



Contributors *

≡ Traci Snowden × + Add contributor



Step 2: Scroll down to find the “Steps to reproduce” field and tool tip

Steps to reproduce

3000 characters left

Additional metadata for Elsevier datasets

Date the data was collected *



Outline how you arrived at your data, to help others understand how the data was gathered and how to reproduce your research. For instance, which methods and protocols were used, and instruments, reagents, software or workflows if any?



Mendeley Data: Case Study 2

Assigning an Embargo Date

[Sign In / Register](#)

DNA-TFAM smFRET

Published: 15 August 2023 | Version 1 | DOI: 10.17632/4whngps32r.1

Contributors: Hyun Huh, Jiayu Shen, Aparna Ramachandran, Yogeeshwar Ajjugal, [Smita Patel](#), [Sang-Hyuk Lee](#)



Embargo: 13 February 2024, 12:00 AM UTC

This dataset will be made public in 153 days

What does under embargo mean?

When publishing a dataset, a user may choose to defer the date at which the data becomes available (for example, so that it is available at the same time as an associated article). This means that the description and files of that dataset are not publicly available until the embargo date is reached. Meanwhile, some other information about the dataset - such as the contributors, title, citation and associated articles become available immediately, prior to the embargo.

[Download calendar event](#)

Latest version

Version 1

Embargo date: 13 February 2024 (153 days)

Published: 15 Aug 2023

DOI: 10.17632/4whngps32r.1

Cite this dataset

Huh, Hyun; Shen, Jiayu; Ramachandran, Aparna; Ajjugal, Yogeeshwar ; Patel, Smita; Lee, Sang-Hyuk (2023), "DNA-TFAM smFRET", Mendeley Data, V1, doi: 10.17632/4whngps32r.1

[Copy to clipboard](#)

Step 1: Click on the “Set embargo” link in the dataset draft

My files

 New Folder  Move  Delete

There are currently no files in your dataset
If you publish, a metadata only dataset will be created

Drag files or folders here

Click here to upload

  Set embargo



Step 2: Click the “Embargo Date” radio button

Select a time period or a calendar date by/on which your dataset will become public.

 [Set embargo / Restrict access](#)

[✕ Remove](#)

 [Set embargo](#)

[✕ Remove](#)

Embargo date

The files will be hidden until the embargo date. Embargos are made public at 12am UTC.

Select time period

3 Months

6 Months

1 Year

or

Select date

dd/mm/yyyy



with specific requesters.



Dataverse

- Restricted Access
- Custom Terms
- Request Access workflow
- NIH
- ReUse:

<https://www.nature.com/articles/s41591-022-01834-y>

https://doi.org/10.7910/DVN/258333, Harvard Dataverse, V10'. There are buttons for 'Access Dataset', 'Contact Owner', and 'Share'. A 'Dataset Metrics' section shows '13,975 Downloads'. A 'Description' section contains the same text as the previous screenshot."/>

...as good scientific practices expect that proper credit is given via citation. Please use the data on this page.

The following Custom Dataset Terms have been defined for this dataset.

You should take precedence over any less restrictive use terms that apply generally to Dataverse.

As part of the Brain Genomics Superstruct Project (GSP) of Harvard University and the Harvard Dataverse, I agree to the following:

I will not share the identity of or attempt to contact any of the included human subjects, nor will I share any of the distributed data to any other data that might contain information about the included human subjects.

In certain circumstances will the code that would link these data to Protected Health Information be given to you. This information about individual human subjects is released to me under these Open Access Data Release Terms and regulations imposed by my institution. This may mean that I need my research to be reviewed by a committee that oversees research on human subjects e.g., my Internal Review Board or Institutional Review Board. Different committees operate under different national, state, and local laws and may interpret regulations differently, so it is important to ask about this.

5. I may redistribute original GSP Open Access data and any derived data as long as the data are redistributed under these same Data Use Terms.

6. I will acknowledge the use of GSP data and data derived from GSP data when publicly presenting any results or algorithms that benefitted from their use.

a. Papers, book chapters, books, posters, oral presentations, and all other printed and digital presentations of results derived from GSP data should contain the following word on in the acknowledgments section: "GSP Data were provided by [name] via the Harvard Dataverse."



Case Study: NIH researcher Using Vivli to access data



The screenshot shows the Vivli website header with navigation links: Home, About, Members, News & Events, Resources, Portals, and a LOG II button. The main content area features the title "Predicting Treatment Response to Tumor Necrosis Factor Inhibitors in Patients with Ankylosing Spondylitis" in large white text on a dark blue background with a network diagram.

Lead Investigator: Michael Ward NIH

Title of Research Proposal: Predicting Treatment Response to Tumor Necrosis Factor Inhibitors in Patients with Ankylosing Spondylitis

Vivli Data Request: 3369

Funding Source: None

Potential Conflicts of Interest: I have no financial or commercial conflicts of interest in the proposed work.

Summary of the Proposed Research:

Axial spondyloarthritis (axial SpA) is a group of inflammatory spine conditions that affects 0.9-1.4% of general population, and ankylosing spondylitis (AS) is the prototypic disease. Tumor Necrosis Factor inhibitors (TNFi) have been widely used as the second line treatment for patients with active AS when patients have inadequate response to non-steroidal anti-inflammatory drugs (NSAIDs) or cannot tolerate NSAIDs. The treatment response to TNFi, however, is heterogenous. In our previous systematic review of randomized control trials of TNFi in patients with AS, about one-half of the participants (39.% to 58.9%) achieved the Assessment in SpondyloArthritis international Society 40%

6 clinical trials accessed from Pfizer and AbbVie

Public Disclosures:

1. Wang R, Dasgupta A, Ward M. Predicting Major Treatment Response to Tumor Necrosis Factor Inhibitors in Patients with Ankylosing Spondylitis [abstract]. *Arthritis Rheumatol.* 2020; 72 (suppl 10).

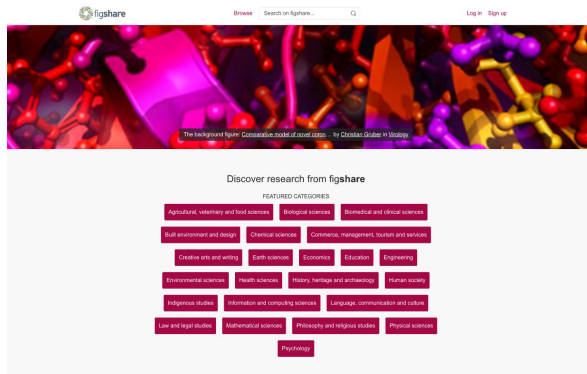
<https://acrabstracts.org/abstract/predicting-major-treatment-response-to-tumor-necrosis-factor-inhibitors-in-patients-with-ankylosing-spondylitis/>.

2. Wang R, Dasgupta A, Ward MM. Predicting Probability of Response to Tumor Necrosis Factor Inhibitors for Individual Patients With Ankylosing Spondylitis. *JAMA Netw Open.* 2022;5(3):e222312. [doi:](https://doi.org/10.1001/jamanetworkopen.2022.2312)

[10.1001/jamanetworkopen.2022.2312](https://doi.org/10.1001/jamanetworkopen.2022.2312)



figshare.com



Discover research from figshare

FEATURED CATEGORIES

- Agricultural, veterinary and food sciences
- Biological sciences
- Biomedical and clinical sciences
- Built environment and design
- Chemical sciences
- Commercial, management, business and services
- Creative city and planning
- Earth sciences
- Economics
- Education
- Engineering
- Environmental sciences
- Health sciences
- History, heritage and archaeology
- Human society
- Informatics studies
- Information and computing sciences
- Language, communication and culture
- Law and legal studies
- Mathematical sciences
- Philosophy and religious studies
- Physical sciences
- Psychology

ALL CATEGORIES SEARCH

sort by: Posted date

JOURNAL CONTRIBUTION

10.1177/1043478X231192160-1
Supplemental material for ...
Posted on 2023-09-15 in SAGE Journals
Xin Yi Lin

COLLECTION
Haplogrouped Natural Killer Cell Therapy as an Adjuvant to Stem Cell Transplantation for ...
Posted on 2023-09-15 in SAGE Journals
Uday Kukarni

10.1177/09638897231198178-1
Supplemental material for ...
Posted on 2023-09-15 in SAGE Journals
Uday Kukarni

10.1177/09638897231198178-2
Supplemental material for ...
Posted on 2023-09-15 in SAGE Journals
Uday Kukarni

10.1177/09638897231198178-3
Supplemental material for ...
Posted on 2023-09-15 in SAGE Journals
Uday Kukarni

10.1177/09638897231198178-4
Supplemental material for ...
Posted on 2023-09-15 in SAGE Journals
Uday Kukarni

10.1177/09638897231198178-5
Supplemental material for ...
Posted on 2023-09-15 in SAGE Journals
Uday Kukarni

10.1177/09638897231198178-6
Supplemental material for ...
Posted on 2023-09-15 in SAGE Journals
Uday Kukarni

Figshare+

store, share, discover

Discover research from Figshare+

87,745 views 153,187 downloads more stats...

ALL CATEGORIES SEARCH

sort by: Posted date

COLLECTION
Supporting data for "Low-coverage whole genome sequencing for a highly selective cohort of seve...
Software posted on 2023-09-12
Renata Santos

SOFTWARE
Source Code for "Low-coverage whole genome sequencing for a highly selective cohort of seve...
Software posted on 2023-09-12
Renata Santos

COLLECTION
Datasets for Deep Imaging Inside Scattering Media Through Virtual Spatio-tempor...
Dataset posted on 2023-08-18
Yiwei Zhang

DATASET
Dataset for Deep Imaging Inside Scattering Media Through Virtual Spatio-tempor...
Dataset posted on 2023-08-18
Yiwei Zhang

COLLECTION
Supporting data for "Low-coverage whole genome sequencing for a highly selective cohort of seve...
Software posted on 2023-09-12
Renata Santos

COLLECTION
Severe COVID-19 Patient Cohort Clinical History
Dataset posted on 2023-08-11
Renata Santos

COLLECTION
FASTQ Alignment and Pre-Processing Quality Control Reports
Dataset posted on 2023-08-11
Renata Santos

COLLECTION
Dataset Validation - Concordance Files
Dataset posted on 2023-08-11
Renata Santos

COLLECTION
Dataset Validation - VCF Files
Dataset posted on 2023-08-11
Renata Santos

COLLECTION
Dataset Validation - CRAM Files
Dataset posted on 2023-08-11
Renata Santos

COLLECTION
Dataset supporting "Inhibitory feedback from the motor circuit gates mechanosensory ...
Dataset posted on 2023-08-09
Sankalp Kumar

COLLECTION
Pathology Images of Scars and Microphages (PLSM)
Dataset posted on 2023-08-07

Figshare for Institutions

National Heart, Lung and Blood Institute

Discover research from National Heart, Lung, and Blood Institute

12,671 views 2,598 downloads more stats...

ALL CATEGORIES SEARCH

sort by: Posted date

DATA SET
Data on phase separation of SigBp3 protein in Vivo
Dataset posted on 2023-08-14
Lisa Greene

DATA SET
Data Supporting AAV-mediated delivery of secreted acid α-glucosidase with enhanced ...
Dataset posted on 2023-07-07
Naveen Kumar Shetty

FIGURE
Three-dimensional structure of pulmonary blood vessels in a post COVID-19 patient
Figure posted on 2023-07-25
Han Han

FIGURE
Data from Petersen et al. The role in mitochondrial free Ca2+ during ischemia - Cell Report...
Figure posted on 2023-07-18
Courtney E. Petersen

DATA SET
Primary data for "TM6SF3 disrupts mitochondrial respiration and may mediate ...
Dataset posted on 2023-07-13
Paul Heung

FIGURE
Data supporting "Methionine sulfoxide reduces and STAR33 constitute a ...
Figure posted on 2023-07-05
Jungmi Lee

FIGURE
Data for paper on High Expression Levels of Hsp104 Cures the Yeast [PSH] Prion ...
Dataset posted on 2023-06-20
Lisa Greene

COLLECTION
Response to oxidative stress of AML12 hepatocyte cells with knockout of methionin...
Posted on 2023-06-09
Jose Abraham Trujillo Hernandez

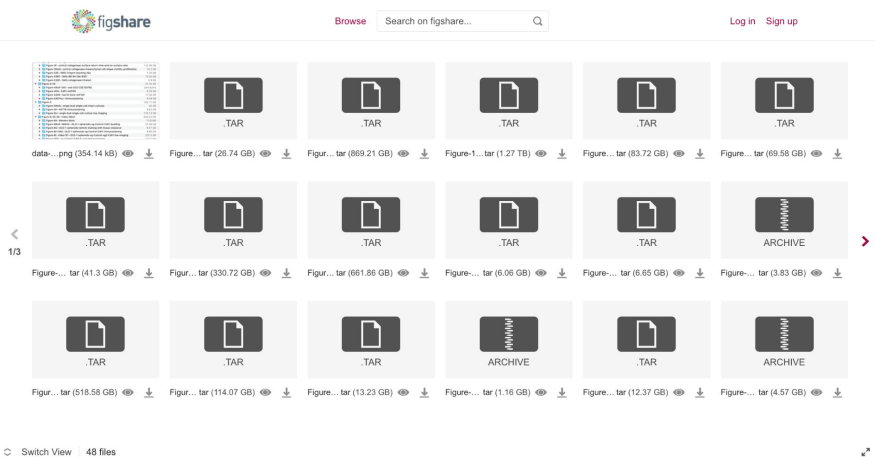
DATA SET
Response to oxidative stress of AML12 hepatocyte cells with knockout of methionin...
Posted on 2023-06-09

DATA SET
Dataset for Reorganization of Mitochondria-Organellar Interactions during Postnatal ...
Dataset posted on 2023-04-14

DATA SET
Data for paper on the binding of Hsp70 to the N-terminal domain of Hsp104
Dataset posted on 2023-04-14

COLLECTION
FRET-CLEM, FLIM, FREM, TIRF, and confocal images





Download all (4.98 TB)

~5TB total

On Figshare+:

- Datasets >20GB up to many TBs
- Files up to 5TB
- FTP and API for file upload
- API to automate files and metadata creation
- File preview for 1,200+ file types
- Upload Zipped and compressed files
- Folder structure with file preview coming 2024!

Switch View 48 files

Budding Epithelial Morphogenesis Driven by Cell-Matrix vs. Cell-Cell Adhesion

[Cite](#) [Download all \(4.98 TB\)](#) [Share](#) [Embed](#) [Collect](#)

Dataset posted on 2021-05-22, 07:01 authored by [Shaoh Wang](#), [Kazuo Matsumoto](#), [Samantha Lish](#), [Alexander Cartagena-Rivera](#), [Kenneth Yamada](#)

USAGE METRICS ?

532 views 113 downloads 0 citations

CATEGORIES

- Evolution of developmental systems

KEYWORDS

branching morphogenesis

budding morphogenesis

epithelial morphogenesis salivary gland

cell-matrix adhesion cell-cell adhesion

differential adhesion E-cadherin

integrin Developmental Biology

LICENCE

CC BY 4.0

FUNDING

Craniofacial Developmental Dynamics

National Institute of Dental and Craniofacial Research

[Find out more...](#)

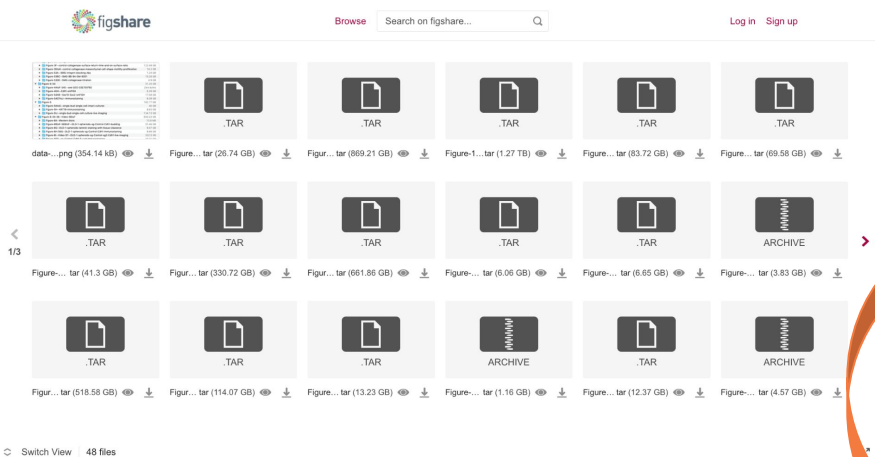
Mechanobiology (ZIAE000094)

Cell Fate and Tissue Dynamics in Salivary Gland Development (K99DE27982)

EXPORTS

<https://doi.org/10.6084/m9.figshare.14582760.v1>





Link to NIH Intramural Project Funding

FUNDING

Craniofacial Developmental Dynamics

National Institute of Dental and Craniofacial Research

[Find out more...](#)

Mechanobiology (ZIAEB000094)

Cell Fate and Tissue Dynamics in Salivary Gland Development (K99DE27982)

Switch View 48 files

Budding Epithelial Morphogenesis Driven by Cell-Matrix vs. Cell-Cell Adhesion

[Cite](#) [Download all \(4.98 TB\)](#) [Share](#) [Embed](#) [Collect](#)

Dataset posted on 2021-05-22, 07:01 authored by [Shaoh Wang](#), Kazuo Matsumoto, Samantha Lish, [Alexander Cartagena-Rivera](#), Kenneth Yamada

USAGE METRICS ?

532 views 113 downloads 0 citations

CATEGORIES

- Evolution of developmental systems

KEYWORDS

branching morphogenesis

budding morphogenesis


epithelial morphogenesis salivary gland

cell-matrix adhesion cell-cell adhesion

differential adhesion E-cadherin

integrin Developmental Biology

LICENCE

 CC BY 4.0

CRANIOFACIAL DEVELOPMENTAL DYNAMICS

National Institute of Dental and Craniofacial Research

[Find out more...](#)

MECHANOBIOLOGY (ZIAEB000094)

CELL FATE AND TISSUE DYNAMICS IN SALIVARY GLAND DEVELOPMENT (K99DE27982)

Dimensions Support Register

Grant

Craniofacial Developmental Dynamics

Funder: National Institute of Dental and Craniofacial Research (NIDCR)
Grant number: ZIAEB000094 - [Original description](#)

Investigators

KENNETH YAMADA - National Institute of Dental and Craniofacial Research
PI

Research organization

National Institute of Dental and Craniofacial Research, United States

Abstract

This project is focused primarily on determining mechanisms of morphogenesis and maintenance of salivary glands and other organs. We are addressing the following major questions: 1. How do embryonic salivary glands and other branched organs generate their characteristic branched architectures during the process of branching morphogenesis? Specifically, how is the formation of clefts, buds, and ducts mediated and coordinated at molecular and biophysical levels? How can we facilitate bioengineering for organ replacement - particularly of salivary glands - by understanding branching morphogenesis and by promoting specific steps of this process? 2. What are the contributions of selective, local regulation of organ-specific gene expression, cell adhesion, embryonic cell sorting, extracellular matrix, integrins, signal

[More](#)

Keywords

3 Dimensional; Address; Adhesives; Architecture; Biomedical Engineering; Biophysics; Bone Marrow; Cell Adhesion; Cell Fraction; Cell Separation; Cell-Cell Adhesion; Cell-Matrix Junction; Cells; Characteristics; Collaborations; Complex; Development; Duct (organ) structure; Embryo; Embryonic Development; Epithelial; Epithelial Cells; Epithelium; Extracellular Matrix; Gene Expression; Gene Expression Profile; Genetic Recombination; Gland; Glands; Human; Integrins; Laboratories; Maintenance; Mediating; Mesenchymal; Molecular; Monoclonal; Morphogenesis; Music; Muscle; Neuronic; Organ; Paired Gland; Process; RNA; Radiation induced damage; Regulation; Salivary; Salivary Glands; Signal Transduction; Specificity; System; Therapeutic; Tissue Engineering; Tissues; Universities; blastomere structure; cell motility; craniofacial; craniofacial development; craniofacial tissue; experimental study; granulocyte; in vivo; molecular marker; novel; organ growth; restoration; transcriptomics

Details

Funding amount
USD 12,608,315

Funding period
2009 - 2022
1 Jan 1 Jan

Program
Intramural Research (Funding Mechanism)

Resulting publications
85

Research categories

Fields of Research (ANZSRC 2020)
31 Biological Sciences
3101 Biochemistry and Cell Biology

<https://doi.org/10.6084/m9.figshare.14582760.v1>



Group data into Collections

Collect related materials together

Both the collection and each item within it are assigned their own DOIs.

This allows researchers the option of using either the collection level DOI to refer to all of the collected datasets, or the item level DOIs to refer to specific items within the collection.

In the THINGS-data collection, NIMH researchers have shared 12 large-scale neuro-imaging and behavioral datasets for the study of natural object representations in the brain and in behavior.

<https://doi.org/10.25452/figshare.plus.c.6161151.v1>

THINGS-data: A multimodal collection of large-scale datasets for investigating object representations in brain and behavior

Follow Posted on 2023-01-17 - 14:30 authored by Martin Hebart

Here we provide all datasets which are part of the THINGS-data collection comprising functional MRI, magnetoencephalographic recordings, and 4.70 million similarity judgments in response to thousands of photographic images for up to 1,854 object concepts. THINGS-data is unique in its breadth of richly-annotated objects, allowing for testing countless hypotheses at scale while assessing the reproducibility of previous findings. Beyond the unique insights promised by each individual dataset, the multimodality of THINGS-data allows combining datasets for a much broader view into object processing than previously possible.

CITE THIS COLLECTION

DataCite

Hebart, Martin; Contier, Oliver; Teichmann, Lina; Rockter, Adam; Zheng, Charles; Kidder, Alexis; et al. (2023). THINGS-data: A multimodal collection of large-scale datasets for investigating object representations in brain and behavior. Figshare+. Collection. <https://doi.org/10.25452/figshare.plus.c.6161151.v1>

Copy citation

<https://doi.org/10.25452/figshare.plus.c.6161151.v1> Copy DOI

or cite all items

Select your citation style and then place your mouse over the citation text to select

FUNDING

Intramural Research Program of the National Institutes of Health (Z)

ERC Starting Grant project COREDIM (101039712)

Research group grant by the Max Planck Society awarded to MNH

Object, face, body and scene representations in the human brain

National Institute of Mental Health

Find out more...

Machine Learning Team

National Institute of Mental Health

Find out more...

REFERENCES

- <https://github.com/ViCCo-Group/THINGS-data>
- <https://doi.org/10.7554/eLife.82580> View PDF

RESEARCH INSTITUTION(S)

National Institute of Mental Health, Max Planck Institute for Human Cogn

CONTACT EMAIL

hebart@cbs.mpg.de

ASSOCIATED PREPRINT DOI

- <https://doi.org/10.7554/eLife.82580>

USAGE METRICS

1625

views

2

citations



Read the associated publication

THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior

figshare+

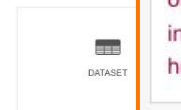
12 results found



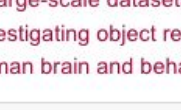
THINGS-data: fMRI PRF AFNI inputs
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: MEG preprocessed dataset
Dataset posted on 2023-01-17 old version
Martin Hebart



THINGS-data: MEG BIDS raw dataset
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: fMRI Single Trial Responses (nifti format)



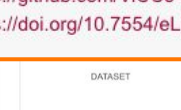
THINGS-data: fMRI BIDS raw dataset
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: fMRI Brain Masks
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: fMRI ICA Noise Regressors
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: fMRI Regions of Interest
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: fMRI Noise Ceilings
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: fMRI Single Trial Responses (table format)
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: fMRI cortical surface flat maps
Dataset posted on 2023-01-17
Martin Hebart



THINGS-data: Behavioral odd-one-out data and code
Dataset posted on 2023-01-17
Martin Hebart

The collection links to the associated [data paper](#) as well as to a [GitHub repository](#) containing scripts and notebooks for reproducing the researchers' analyses.

Read the associated publication

THINGS-data, a multimodal collection of large-scale datasets for investigating object representations in human brain and behavior

REFERENCES

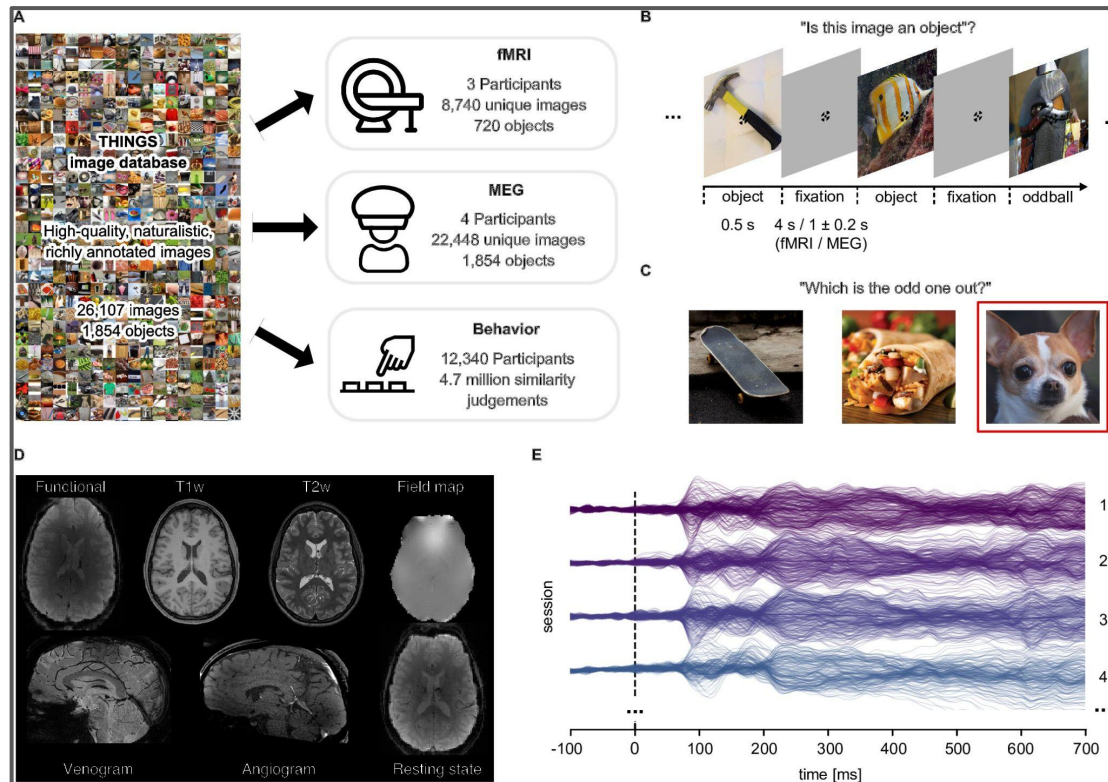
- <https://github.com/ViCCo-Group/THINGS-data>
- <https://doi.org/10.7554/eLife.82580> View PDF



Multi-modal Dataset Designed for Reuse

The associated [eLifeSciences data paper](#) provides the following overview of the complete data collection, as illustrated by the figure at right:

- (A) THINGS-data comprises MEG, fMRI and behavioral responses to large samples of object images taken from the THINGS database.
- (B) In the fMRI and MEG experiment, participants viewed object images while performing an oddball detection task (synthetic image).
- (C) The behavioral dataset comprises human similarity judgements from an odd-one-out task where participants chose the most dissimilar object amongst three options.
- (D) The fMRI dataset contains extensive additional imaging data.
- (E) The MEG dataset provides high temporal resolution of neural response measurements in 272 channels. The butterfly plot shows the mean stimulus-locked response in each channel for four example sessions in one of the participants.



Use of multiple repositories with Figshare

The THINGS-data collection was designed for reuse and extends beyond the Figshare collection alone, making use of additional open discipline specific and generalist repositories, to best share the complete data collection.

The data availability statement explains where each part of the data collection can be found:

- OpenNeuro
- Figshare
- OSF

Data availability

All parts of the THINGS-data collection are freely available on scientific data repositories. We provide the raw MRI (<https://doi.org/10.18112/openneuro.ds004192.v1.0.5>) and raw MEG (<https://doi.org/10.18112/openneuro.ds004212.v2.0.0>) datasets in BIDS format (Gorgolewski et al., 2016) on OpenNeuro (Markiewicz et al., 2021). In addition to these raw datasets, we provide the raw and preprocessed MEG data as well as the raw and derivative MRI data on Figshare (Thelwall and Kousha, 2016) at <https://doi.org/10.25452/figshare.plus.c.6161151>. The MEG data derivatives include preprocessed and epoched data that are compatible with MNE-python and CoSMoMVPA in MATLAB. The MRI data derivatives include single trial response estimates, category-selective and retinotopic regions of interest, cortical flatmaps, independent component based noise regressors, voxel-wise noise ceilings, and estimates of subject specific retinotopic parameters. In addition, we included the preprocessed and epoched eyetracking data that were recorded during the MEG experiment in the OpenNeuro repository. The behavioral triplet odd-one-out dataset can be accessed on OSF (<https://osf.io/f5rn6/>).



Connect with the Repositories

Dataverse: [NIH-DMP Guidance for Harvard Dataverse](#) and support@dataverse.harvard.edu

Dryad: [Dryad submission requirements](#); [Good Data Practices](#); [Dryad's role in the NIH's new Policy for Data Management and Sharing](#); [E hello@datadryad.org](mailto:hello@datadryad.org)

Figshare: [Guide to sharing NIH-funded research](#) and [How to write a Data Management Plan \(DMP\) and include Figshare in your data sharing plans](#) [E info@figshare.com](mailto:info@figshare.com)

OSF: [Common questions and support documentation](#) and [Creating a data management plan \(DMP\) document](#)

Mendeley Data: [Elsevier sharing policy](#), [Guidance on publishing for researchers](#), and [Elsevier journal data guidelines](#)

Vivli: [Data submission checklist](#)

Zenodo: [NIH Data Management and Sharing Plan Guidance](#) and info@zenodo.org



Connect with GREI

- Join the [GREI Google Group](#) to receive updates on GREI activities. ADD EMAIL
- Read the latest posts on the [GREI Community blog](#).
- Browse the resources available in the [GREI Community](#) on Zenodo.
- Check out the [GREI Training & Outreach Calendar](#) for information about upcoming events from the individual GREI repositories.
- Help us engage with *your* research communities with resources, a custom presentation, or communications.



Your Feedback

~

What do you need from generalist repositories?



Questions?

