



GREI Generalist Repository
Ecosystem Initiative

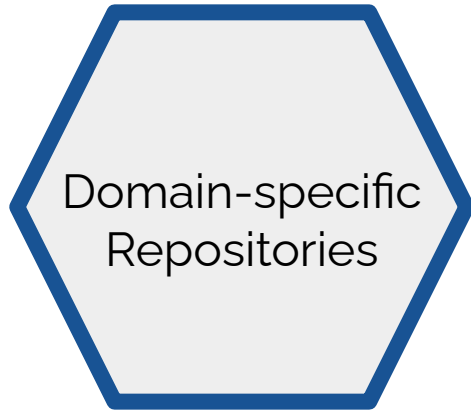
GREI Collaborative Workshop

Guidance for sharing data in
generalist repositories

2:15-3:45pm Eastern

January 24, 2023

NIH Research Data Ecosystem



Hi! We're your hosts for today



David Scherer

Customer Consultant, Pure and
Research Data Management
Elsevier



Jess Herzog

Head of Publishing Services
Dryad



Here's what we'll cover

- Guide to data sharing (20 min)
- How to generate complete, high-quality metadata for submission in a generalist repository (30 min)
- Breakout sessions (30 min): facilitated session for questions, feedback, etc.
- Wrap-up (5-10 min)



POLL

Anonymous



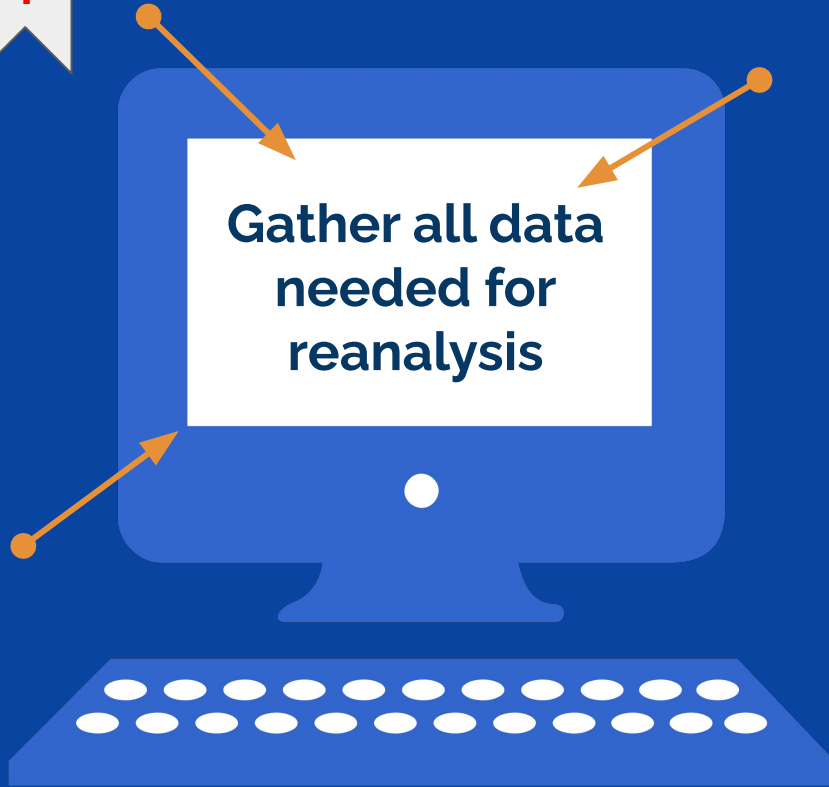
Have you shared data or supported a
researcher in sharing data?



~~Survival~~ GUIDE for data sharing



1



Consider including:

- 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-CCo)
- ✓ Content from published articles, grants, or patents
- ✓ Data from 3rd party with restricted terms-of-use
- ✓ Identifiable human subjects data
- ✓ Location information for endangered and/or vulnerable species



Choose **open** file formats



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



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 no 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

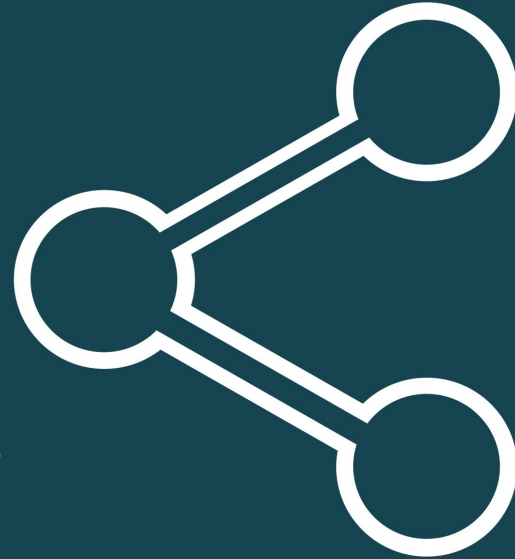
Common:
Core Metadata
Persistent Identifiers
Discoverable
Flexibility
Open access, FAIR
Metrics

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



Elements of high-quality metadata for a **generalist repository**

30 minutes

Data



Metadata



<https://dataedo.com/kb/data-glossary/what-is-metadata>



zenodo



Let's meet our presenters!



Sonia Barbosa

Manager of
Curation,
Harvard
Dataverse



Gretchen Gueguen

Product Owner,
Center for Open
Science (OSF)



Julie Wood

Senior Director,
Vivli



Sara Gonzales

Senior Data Librarian,
Northwestern
University,
Zenodo



**Ana Van Gulick,
PhD**

Government and
Funder Lead,
Figshare



Dryad is



An **open data publishing platform & community** committed to the **open availability** and **routine re-use** of all research data

49,600+ data publications

193,300+ researchers

69,100+ international institutions

1,270+ academic journals





Dryad submission form

(page 1)

DOI assigned upon submission & does not change

DRYAD Admin | Explore data | About | Help | My datasets | Logout

[Describe dataset](#) [Upload files](#) [Review and submit](#)

Describe your dataset

Preliminary information

My data is related to:

a manuscript in progress a published article other or not applicable

Please provide the following information. You may either enter the information and leave it or choose to autofill your dataset based on the information you supply below.

Journal Name: * Manuscript Number *

[Import Manuscript Metadata](#)

Dataset: Basic information

Dataset Title *

Author(s)

First Name *	Last Name *	Institutional Affiliation *	Author Email *
<input type="text" value="Jess"/>	<input type="text" value="Herzog"/>	<input type="text" value="Bioengineering Center"/>	<input type="text" value="jherzog@datadryad.org"/>

<https://sandbox.orcid.org/0000-0003-0361-6671> Corresponding Author

[+ Add author](#)

Research Domain * Research Facility:

Funding

Granting Organization: * Award Number

[+ Add another funder](#)

Abstract *

Journal Info

Dataset Title

Author(s)

Research Area

Funding Info

Abstract



Dryad submission form

(page 1)

Keywords: Adding keywords improves the findability of your dataset. E.g. scientific names, method type

Analytical Framework

Methods: How was this dataset collected? How has it been processed?

Paragraph **B** *I*

P

Press Alt 0 or Option 0 for help using the rich text editor with keyboard only.

Usage Notes: What programs and/or software are required to open the data files included with your submission? If proprietary, include open-source alternatives.

Paragraph **B** *I*

P

Press Alt 0 or Option 0 for help using the rich text editor with keyboard only.

Related works

Are there any preprints, articles, datasets, software packages, or supplemental information that have resulted from or are related to this Data Publication?

Work Type	Identifier or external url
<input checked="" type="checkbox"/> Article	<input type="text" value="example: https://doi.org/10.1594/PANGAEA.726855"/> remove
<input type="checkbox"/> Dataset	
<input type="checkbox"/> Preprint	
<input type="checkbox"/> Software	
<input type="checkbox"/> Supplemental Information	
<input type="checkbox"/> Data Management Plan	

[Privacy policy](#) [Accessibility policy](#) [Terms of service](#) Version: v0.8.0.3;

[Contact us](#) [Follow us on Twitter](#) [Check out our blog](#)

Copyright (c) 2023 Dryad

Keywords

Methods

Usage Notes

Related Works

Dryad submission form (page 2)

- **CCo license** for Dryad files; **CCBY** for Zenodo
- **Open, accessible file types** (non-proprietary)
- **Tabular data validator** automatically checks the format and structure of CSV, XLS, XLSX file formats
- 300GB limit, up to 1TB

DRYAD Admin | Explore data | About | Help | My datasets | Logout

Describe dataset | Upload files | Review and submit

Upload your files

You may upload data via two mechanisms: directly from your computer, or from a URL on an external server (e.g., Box, Dropbox, AWS, lab server). We do not recommend using Google Drive.

We require that you include a [README.md file based on our template](#) in order to provide key information for understanding and reuse of your data.

If you prefer, you can edit the Markdown online at [hackmd.io](#).

- Open the link (above)
- Create a new note by clicking "+" at the top left of the screen
- Copy and paste your text
- Click the three dots on the top right to download as a Markdown file
- Make sure your filename is 'README.md', and upload the file under the *Data* category

Software and Supplemental Information can be uploaded for publication at [Zenodo](#). You will have the opportunity to choose a separate license for your software on the review page.

DRYAD Data

Required: README.md
e.g., csv, fasta

Choose files

Enter URLs

zenodo Software

e.g., code packages, scripts

Choose files

Enter URLs

zenodo Supplemental information

e.g., figures, supporting tables

Choose files

Enter URLs

Files

Filename	Status	Tabular data check	URL	Type	Size	Actions
Dryad_README_templatev2.txt	Uploaded			Data	1.82 KB	Remove

Back to Describe dataset | Proceed to review

README file

1 data file



Dryad submission form

(page 3)

Private for Peer
Review option

Related Works

Article: <https://doi.org/10.1016/j.mex.2022.10164>

◀ Edit Description

Review files



Data files hosted by Dryad

[Fortuny-et-al-2021_Tomato-fruit-quality-traits-and-metabolite-content-are-affected-by-reciprocal-effect-and-heterosis_RawData.xls](#) 3.32 MB

[README.txt](#) 2.07 kB

◀ Edit Files

Private for peer review

By choosing this option, your dataset will be private during your related article's peer review period. You will have access to a private dataset download URL to be shared with collaborators or the journal. Your dataset will not enter curation or be published. Because we may not have the status of your related article, the default for this period is six months. Please [email us](#) or uncheck this box at any point if your dataset is ready to enter curation.

Keep my dataset private while my related article is in peer review

Agree to terms

License and terms of service for data

- By checking this box, I agree to the license [CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](#) *
- By checking this box, I agree to [Dryad's Terms of Service](#) *



Payment

Dryad charges a fee for data publication that covers curation and preservation of published datasets. Upon publication of your dataset, you will receive an invoice for \$120 USD. We're sensitive to the fact that fees for individual researchers can be a burden and create inequities. If you'd like to request a fee waiver, please contact help@datadryad.org.

- I agree to Dryad's payment terms

◀ Back to Upload

Submit

Dryad dataset view

- Track **metrics**: citations, views, & downloads
- **Shareable URL** allows access to all data files (useful for journal peer-review process)

<https://doi.org/10.7959/dryad.w6m905qk8>

Citation

Wang, Xiao-Yan, Song, Gao, Wang, Jiang, Tu, Fei-Hai (2023), Interactions between soil microbes and native species drive a diversity-invasibility relationship, Dryad, Dataset, <https://doi.org/10.7959/dryad.w6m905qk8>

Abstract

Soil microbes can affect both the invasiveness of exotic plants and the invasibility of native plant communities, but it still remains unclear whether soil microbes can influence the relationship between native plant species diversity and community invasibility. We constructed native plant communities with three levels of species richness (one, three, or six species) in un-sterilized or sterilized soil (i.e., with or without soil microbes) and let them not be invaded by exotic plant species or invaded by one of three exotic species (*Solidago canadensis*, *Erigeron canadensis* or *Symphyotrichum subulatum*) highly invasive in China. The soils conditioned by the native plant communities not invaded by the exotic species were used as soil microbe inocula to test whether richness-induced differences in soil microbes affect the growth of each of the three invasive species. Compared with the presence of soil microbes, the absence of soil microbes weakened the negative species richness-invasibility relationship, indicating that soil microbes can contribute to the higher invasion resistance of more diverse native plant communities. In the presence of soil microbes, the higher invasion resistance of more diverse communities was mainly ascribed to the complementarity effect. However, soil microbes from communities with a higher species richness did not have a stronger negative effect on the growth of any of the three invasive species. We conclude that soil microbes can alter the diversity-invasibility relationship through promoting the complementarity effect on the community invasion resistance. Our results highlight the importance to integrate the role of soil microbes in testing the diversity-invasibility hypothesis.

Usage Notes

Exp1: Colume D-M for plot with invasive species *Solidago canadensis*(SC)

Colume N-W for plot with invasive species *Erigeron canadensis*(EC)

Colume X-G for plot with invasive species *Symphyotrichum subulatum*(SS)

Exp2: Colume AH-AO

Reviewer URL

You can share this version of your dataset files with others using the url below.

Click the *Select Text* button and then copy and paste the URL.

Sharing Link

<https://dryad-stg.cdlib.org/stash/share/mQr8ihkmlb6jWfniMWUSkeTEHtKUpEEWeopJ5kZZKA>

Select Text

Metrics

 0 views

 0 downloads

 0 citations

Keywords

Biological sciences

phylogenetic distance

License

This work is licensed under a [CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication license](https://creativecommons.org/licenses/by/4.0/).



The Dataverse Project

Open source research data
repository software

95 Installations around the
world

DATVERSE REPOSITORIES - A WORLD VIEW

95 Installations



Researchers

Enjoy full control over your data. Receive *web visibility, academic credit, and increased citation counts*. A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research program, makes your data more discoverable to the research community, and satisfies data management plans. [Want to set up your personal Dataverse collection?](#)



Journals

Seamlessly manage the submission, review, and publication of data associated with published articles. Establish an *unbreakable link* between *articles in your journal* and *associated data*. Participate in the open data movement by using a Dataverse collection as part of your journal data policy or list of repository recommendations. [Want to find out more about journal Dataverse collections?](#)



Institutions

Establish a research data management solution for your community. Federate with a growing list of Dataverse repositories worldwide for increased discoverability of your community's data. Participate in the drive to set norms for sharing, preserving, citing, exploring, and analyzing research data. [Want to install a Dataverse repository?](#)



Developers

Participate in a vibrant and growing community that is helping to drive the norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization and exploration tools*, or other research and data archival systems with the Dataverse Project. [Want to contribute?](#)

The Harvard Dataverse Repository: Metadata and Files

Organize datasets and gather metrics in your own repository.

A dataverse is a container for all your datasets, files, and metadata.

[Add a dataverse +](#)

Deposit and share your data. Get academic credit.

Harvard Dataverse is a repository for research data. Deposit data and code here.

[Add a dataset +](#)

Publishing your data is easy on Harvard Dataverse!

Learn about getting started creating your own dataverse repository here.

[Getting started !\[\]\(d0262bbe9d2356661a2e89321dfcc781_img.jpg\)](#)

- **Dataverse Collection**

- **Datasets**

- **Files, Metadata, Terms, and Versions**

Published content:

[Dataverses \(6,117\)](#)

[Datasets \(160,541\)](#)

[Files \(1,880,729\)](#)

Access type:

[Public \(1,827,599\)](#)

[Restricted \(52,943\)](#)

[Embargoed then Public \(173\)](#)

[Embargoed then Restricted \(14\)](#)

Dataverse Category

[Research Project \(2,266\)](#)

[Researcher \(1,912\)](#)

[Organization or Institution \(471\)](#)

[Research Group \(431\)](#)

[Journal \(127\)](#)

[Laboratory \(93\)](#)

[Department \(69\)](#)

[Teaching Course \(44\)](#)

The Harvard Dataverse Repository: Metadata and Files

Dataverse Collection creation

- **Citation metadata** always required
- **Multiple metadata blocks** selection
- Make metadata fields required, optional, **template feature** for multiple depositors

Dataverse Name * ⓘ Harvard Dataverse	Affiliation ⓘ
Identifier ⓘ https://dataverse.harvard.edu/dataverse/ harvard	Storage ⓘ s3 (Default)
Category * ⓘ Organization or Institution	Description ⓘ This field supports only certain HTML tags.
Email * ⓘ <input type="text"/>	

Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse.

- Citation Metadata (Required) [\[+\] View fields + set as hidden, required, or optional](#)
- Geospatial Metadata [\[+\] View fields](#)
- Social Science and Humanities Metadata [\[+\] View fields](#)
- Astronomy and Astrophysics Metadata [\[+\] View fields](#)
- Life Sciences Metadata [\[+\] View fields](#)
- Journal Metadata [\[+\] View fields](#)
- Computational Workflow Metadata [\[+\] View fields](#)

The screenshot shows the Harvard Dataverse website interface for the Brain Genomics Superstruct Project (GSP) dataset. The page header includes the project logo and the text "NEUROINFORMATICS RESEARCH GROUP". The main content area displays the dataset title "Brain Genomics Superstruct Project (GSP) Dataverse" and its affiliation "(Harvard University)". Below this, there is a description of the dataset, which is a large-scale imaging data set for neuroimaging, behavior, cognitive, and personality data. A warning icon and text prompt users to provide an academic or research institution email address when applying for access. The page also features a search bar, a "Databases" filter, and a list of datasets, with the current dataset being the "Brain Genomics Superstruct Project (GSP)" dataset, published on May 15, 2015, by Buckner, Randy L.; Roffman, Joshua L.; Smoller, Jordan W., 2014.

The Harvard Dataverse Repository: Metadata and Files

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

[Cite Dataset](#) ▾

[Learn about Data Citation Standards.](#)

Dataset deposit

- DOIs(d)
- Publication year(d)
- Title(r)
- Author(r)
- Contact(r)
- Subject(r)
- Description(r)
- Terms(r)
- Versions(d)

- Related publication and identifier

Dataset citation

- Affiliation
- Identifier Type

Citation Metadata ▲

Title * ⓘ

Author * ⓘ

Name * ⓘ **Affiliation** ⓘ +

1) Family Name, Given Name or 2) Orgar Organization XYZ

Identifier Type ⓘ **Identifier** ⓘ

Select... Select...

Point of Contact * ⓘ

Name ⓘ **Affiliation** ⓘ +

1) FamilyName, GivenName or 2) Organi: Organization XYZ

E-mail * ⓘ

name@email.xyz

Description * ⓘ

This field supports only certain HTML tags.

Text * ⓘ +

Related Publication ⓘ

Kaufman, Aaron R., Gary King, and Mayya Komisarchik. 2021. "How to Measure Legislative District Compactness if You Only Know It When You See It." *American Journal of Political Science* 65 (3): 533-50. doi: [10.1111/ajps.12603](https://doi.org/10.1111/ajps.12603)

The Harvard Dataverse Repository: Metadata and Files

Files

- **Tabular data ingest**, with online data analysis tool: **metadata extraction**
- BagIT
- **Folder preservation** preservation
- **Provenance**
- File Types
- File Tag
- **Embargo, access restrictions**

Datasets (78)
 Files (1,974)

Publication Year

2021 (114)
 2020 (14)
 2019 (462)
 2018 (26)
 2017 (21)

[More...](#)

File Type

Text (715)
 Tabular Data (601)
 Data (267)
 Archive (107)
 Code (101)

File Tag

2. Data (141)
 3. Supplementary Program Files (12)
 3. US-Female Data (101)
 1. Documentation (90)
 3. Supplementary Files (37)

[More...](#)

Access

Public (1,969)
 Restricted (5)

File level Metadata extraction

MiinerButhe2.tab
 Aug 4, 2014 - Replication data for: How Robust Standard Errors Expose Methodological Problems They Do Not Fix, and What to Do About It
 Tabular Data - 1.2 MB - 67 Variables, 2673 Observations - UNF:5/jhz...iLw==

2. ReplicationFolder-Data

main_catalist_file.tab
 Sep 4, 2021 - Replication Data for: Statistically Valid Inferences from Differentially Private Data Releases, with Application to the Facebook URLs Dataset
 Tabular Data - 411.5 KB - 41 Variables, 2575 Observations - UNF:6/qVAL...4xg==

	country	ccode	id	year	t	coup	coup_success	coup_fail	latent_personalism	pers_2pl
1	Nigeria	475.0	192	1994.0	0.0	0.0	0.0	0.0	0.4196175	4.665E-4
2	Nigeria	475.0	192	1995.0	1.0	0.0	0.0	0.0	0.4196175	4.665E-4
3	Nigeria	475.0	192	1996.0	2.0	0.0	0.0	0.0	0.5845026	0.5198256
4	Nigeria	475.0	192	1997.0	3.0	0.0	0.0	0.0	0.5845026	0.5198256
5	Nigeria	475.0	192	1998.0	4.0	0.0	0.0	0.0	0.5845026	0.5198256
6	Sudan	625.0	234	1959.0	0.0	1.0	0.0	1.0	0.0	-1.321256
7	Sudan	625.0	234	1959.0	1.0	1.0	0.0	1.0	0.0	-1.321256
8	Sudan	625.0	234	1960.0	2.0	0.0	0.0	0.0	0.0	-1.321256
9	Sudan	625.0	234	1961.0	3.0	0.0	0.0	0.0	0.0	-1.321256
10	Sudan	625.0	234	1962.0	4.0	0.0	0.0	0.0	0.0	-1.321256
11	Sudan	625.0	234	1963.0	5.0	0.0	0.0	0.0	0.0	-1.321256
12	Sudan	625.0	234	1964.0	6.0	0.0	0.0	0.0	0.0	-1.321256

Data explorations/analysis

Current Features



Installations



Collections



Datasets



Files

NIH OTA Adds...



Datasets

- Flexible **Biomedical Metadata** Support through External Vocabularies and Data Dictionaries (UMLS, CEDAR, MeSH)



Files

- Additional Metadata for **Code** files ([Codemeta](#))
 - Support for **Replication Packages** through Workflows and Containers
 - **Sensitive Data** Support through Differential Privacy ([OpenDP](#), [DataTags](#), [PSIprivacy](#))
 - Encryption
-
- Usage Metrics
 - UX/UI Enhancements
 - Interoperability
 - Metadata Harvesting
 - Curation Services
 - [Training and Outreach](#)



Vivli: For sharing and Accessing Clinical Research Data



Upload
Anonymized
datasets (IPD)
from
completed
clinical trials



ARCHIVE

Archive datasets for reuse
Assign contributor roles to
team members (linked to
their ORCID) to allow
downstream CREDIT



ACCESS

Managed Access:
Requesters submit a
proposal and sign DUA;
Key metrics tracked to
generate use reports for
contributors



ANALYZE

Data is reused
and citations are tracked

A Multi-centre, Adaptive, Randomized, Open-label, Controlled Clinical Trial of the Safety and Efficacy of Investigational Therapeutics for the Treatment of COVID-19 in Hospitalized Patients (CATCO: Canadian Treatments for COVID-19), in Conjunction With the Public Health Emergency SOLIDARITY Trial (World Health Organization) (CATCO)

[Study Details](#)[Study Documents](#)[Administrative Details](#)[Usage](#)

Phase

Phase 3

Condition or Disease

COVID-19

Intervention/treatment

Artesunate, Imatinib, Infliximab, ARBs, Dexamethasone, LSALT Peptide

Artesunate, Imatinib, Infliximab, ARBs,
Dexamethasone, LSALT Peptide

Brief Summary

This study is an adaptive, randomized, open-label, controlled clinical trial, in collaboration with countries around the world through the World Health Organization.

Ages Eligible For Study

18 Years and older

Sexes Eligible For Study

All

Accepts Healthy Volunteers

No

Actual Enrollment

Not available

Locations

Canada (57)

Additional Information

This data package is available for download. Vivli will conduct an accelerated review of any requests for this study package.

Primary Registry Name

ClinicalTrials.gov

Primary Registry ID

NCT04330690

Primary Registry Url

<https://clinicaltrials.gov/show/NCT04330690>

STUDY DESIGN

Vivli: What's included in a data package?

Item	Description
<i>Recommended Data Package Set</i>	
Study Protocol	Final protocol with all amendments
Data dictionary	Detailed descriptions of each variable in the dataset, including the definition, source, coding, etc. of the variable
Statistical Analysis Plan	Description of the principal features of the analysis described in the protocol
IPD dataset	Final cleaned individual participant-level data, de-identified/anonymized
Anonymization Guidance	Outlines the method used to anonymize the data
<i>Optional</i>	
Analytic code	Software code used to carry out prespecified and additional analyses
Analysis ready IPD data set	Dataset in a format used to carry out a sponsor's analyses
Case report forms	Forms used to collect the data that is described in the protocol for each trial participant

NOTE: This is a subset of the entire full data package and includes the data that underlies the publication findings (tables, figures)



Upload

- All file formats accepted
- 50GB per deposit
- GitHub integration

Publish

- Data citation
- Control access levels
- Usage statistics / metrics
- DOI versioning
- File previews
- Link to related research

Describe

- DataCite Metadata Schema
- Custom metadata fields
- Funder integration
- Communities feature
- Wide license selection

Compliance with sharing mandates

- Deposit data files and supporting materials
- Restricted files / embargo
- Private record sharing

Persistent Metadata

- DOIs for all deposits
- Reserve DOI
- PID-enabled metadata fields: ORCID, subject vocabularies
- FAIR data

Common Formats

- DataCite
- OAI-PMH
- DublinCore
- JSON
- Schema.org
- REST API
- CSL



Upload

50GB* for each dataset
All file formats accepted

The screenshot shows the 'New upload' page. At the top, there are buttons for 'Delete', 'Save', and 'Publish'. Below this is a section for 'Files' with a 'Choose files' button and a 'Start upload' button. A table lists four files: '2020-06-report-3.json.gz' (5 Mb), 'broker-2019-01-17T13:48:21.bk' (200 Mb), 'pubmed-events.json' (957 Kb), and 'storage_growth.py' (703 B). A note at the bottom states: 'Note: File addition, removal or modification are not allowed after you have published your upload. This is because a Digital Object Identifier (DOI) is registered with DataCite for each upload. (minimum 1 file required, max 50 GB per dataset - contact us for larger datasets)'. The 'Publish' button is highlighted in blue.

Describe

Rich but flexible metadata
Based on DataCite schema
Reserve DOI before

The screenshot shows the 'Describe' page. At the top, there are icons for various upload types: Publication, Poster, Presentation, Dataset, Image, Video/Audio, Software, Lesson, and Other. Below this is the 'Basic information' section, which is required. It includes a 'Digital Object Identifier' field with the value '10.5072/zenodo.682186'. There is an optional note: 'Optional: Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.' There is a 'Reserve DOI' checkbox which is checked. The 'Publication date' is set to '2020-10-15'. The 'Title' is 'Data and descriptions for XYZ research project'. The 'Authors' section lists 'Alice' and 'Smith' with ORCID iDs, and 'Alex' and 'Ioannidis' with ORCID iDs. There is a 'Description' section with a text area containing 'This is to the datasets and their description/documentation for the work published on project XYZ.' The 'Publish' button is highlighted in blue.

Publish

Citable DOI
Export formats

The screenshot shows the 'OpenAIRE Covid-19 publications, datasets, software and projects metadata' page. The page is dated 'August 12, 2020'. It shows '4,620 views' and '438 downloads'. The 'OpenAIRE' logo is prominently displayed. The 'Publication date' is 'August 12, 2020'. The 'DOI' is '10.5072/zenodo.3980491'. The 'Communities' section lists 'OpenAIRE' and 'OpenAIRE Research Graph Zenodo'. The 'License (or files)' section lists 'Creative Commons Zero v1.0 Universal'. The 'Versions' section shows 'Version 1.0' with a date of 'Aug 12, 2020'. The 'Files' section shows 'COVID-19.json.gz' (52.8 MB) and 'mdds27ab78b997c2281146ea643c3d5f6990'. The 'Citations' section shows 'Show only Literature (0) Dataset (0) Software (0) Unknown (0) Citations to this version'. The 'Publish' button is highlighted in blue.

DOI versioning, Usage statistics

Zenodo now supports DOI versioning!

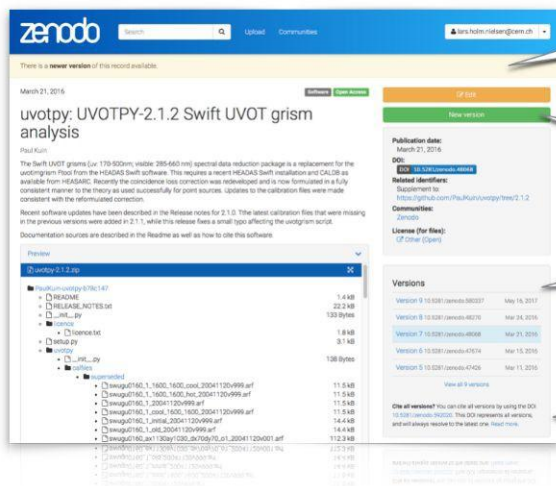
by  Lars Holm Nielsen on May 30, 2017

We are pleased to announce the launch of DOI versioning support in Zenodo - the open research repository from OpenAIRE and CERN. This new feature enables users to update the record's files after they have been made public and researchers to easily cite either specific versions of a record or to cite, via a top-level DOI, all the versions of a record.

DOI versioning support was one of our most requested features for Zenodo, and it has been co-developed by OpenAIRE's Zenodo team and EUDAT's BZSHARE team as an extension module for CERN's Invenio digital repository platform, which powers both Zenodo and BZSHARE.

This update comes hot on the heels of the recent relaunch which made Zenodo faster, improved GitHub integration, integrated support for Horizon 2020 grant information, and enabled 50 gigabyte uploads!

Read more about the inner workings of new feature in the DOI Versioning FAQ:



The screenshot shows the Zenodo record page for 'uvotpy: UVOTPY-2.1.2 Swift UVOT grism analysis'. The page includes a search bar, user profile, and a 'new version' alert. The version history table is as follows:

Version	Date	
Version 9	10.5281/zenodo.580227	May 16, 2017
Version 8	10.5281/zenodo.48276	Mar 24, 2016
Version 7	10.5281/zenodo.48048	Mar 21, 2016
Version 6	10.5281/zenodo.43124	Mar 15, 2016
Version 5	10.5281/zenodo.41405	Mar 11, 2016

The file list includes various calibration files and software binaries, such as 'uvotpy', 'uvotpy_2.1.2.zip', and 'uvotpy_2.1.2.tar.gz'.

Alert if newer version is available

Create a new version

Browse the version history

Cite specific version or the concept representing all versions

4,242

views

3,929

downloads

See more details...

	All versions	This version
Views	4,242	2,854
Downloads	3,929	3,816
Data volume	5.0 TB	4.8 TB
Unique views	4,074	2,793
Unique downloads	337	258

More info on how stats are collected.

Versions

Version v0.9.0 Jul 16, 2018
10.5281/zenodo.1313201

Version v0.8.1 Sep 3, 2017
10.5281/zenodo.883859

Version v0.8.0 Jul 8, 2017
10.5281/zenodo.824567

Version v0.7.1 Jun 5, 2016
10.5281/zenodo.54844

Version v0.7.0 Jan 25, 2016
10.5281/zenodo.45133

[View all 8 versions](#)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.592845](https://doi.org/10.5281/zenodo.592845). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

OSF and Supporting Research Across the Lifecycle

COS Mission: to increase openness, integrity, and reproducibility of research

COS Infrastructure Vision: To empower communities, institutions, and funders to advance rigor and transparency of research.



A tool for research...

The image is a collage of three overlapping screenshots from the Open Science Framework (OSF) ecosystem, demonstrating the process of preregistration and publication of research.

Top Left Screenshot (OSF REGISTRIES): Shows the "Preregistration of Pre..." page. The navigation bar includes "Add New", "My Registrations", "Help", and "Donate". A sidebar on the left lists "Public registration" and "Updates". The main content area is titled "Study Information" and includes sections for "Overview", "Files", "Resources", "Wiki", "Components", "Links", "Analytics", and "Comments". The "Open practice resources" section lists items like "Data", "Analytic code", "Materials", "Papers", and "Supplements".

Top Middle Screenshot (OSFHOMe): Shows the "Preregistration of Preregistration evalu..." page. The navigation bar includes "My Projects", "Search", "Support", "Donate", and "Gretchen Gueguen". The main content area is titled "Preregistration evaluation 2016 / Preregistration of Preregistration evaluation 2016". It lists contributors: Coosje Lisabet Sterre Veldkamp, David Thomas Mellor, Marjan Bakker, Elise Anne Victoire Crompvoets, Courtney K. Soderberg. It also shows the date created (2016-09-02 10:08 AM) and last updated (2017-05-31 04:06 AM). The category is "Uncategorized". A "Wiki" section contains the text: "This component contains the full preregistration document of our study 'Preregistration Evaluation 2016'." A "Files" section lists various files, including "Preregistration of Preregistration evaluation 20...", "Confirmatory_Analyses_and_Follow_up_A...", "Creation_Coding_Scheme.R", "Fig1_preregistration.png", "Preregistration of Study Researcher Degr...", "Protocol Coding Preregistrations.pdf", and "Random_Pre_selection_SPR.R".

Top Right Screenshot (PsyArXiv Preprints): Shows the "Ensuring the quality and specificity of preregistrations" preprint page. The navigation bar includes "My Preprints", "Submit a Preprint", "Search", "Donate", and "Gretchen Gueguen". The main content area is titled "Ensuring the quality and specificity of preregistrations". It lists authors: Marjan Bakker, Coosje Lisabet Sterre Veldkamp, Marcel A.L.M. van Assen, Elise Anne Victoire Crompvoets, How Hwee Ong, Brian A. Nosek, Courtney K. Soderberg, David Thomas Mellor, Jelte Wicherts. It also shows "AUTHOR ASSERTIONS" for "Conflict of Interest: Yes", "Public Data: Available", and "Preregistration: Available".

Bottom Screenshot: Shows the "Ensuring the quality and specificity of preregistrations" preprint page in a browser window. The page includes a "Download" button, a "plaudit" logo, and a "Robert Heineke has endorsed this work." badge. The abstract states: "Researchers face many, often seemingly arbitrary choices in formulating hypotheses, designing protocols, collecting data, analyzing data, and reporting results. Opportunistic use of 'researcher degrees of freedom' aimed at obtaining statistical significance increases the likelihood of obtaining and publishing false positive results and ...". The page also includes a "Supplemental Materials" section and a "Preprint DOI" (10.31234/osf.io/cdgyh).

Data Repository

OSF: Trusted Repository for Data

[Desirable Characteristics for Data Repositories](#) supporting data sharing policies across the research lifecycle

- [NIH first to come out with the characteristics](#) in advance of their Data Management and Sharing policy
- OSF on list of generalist repositories for NIH data, and member of [NIH GREI project](#)
- [USGS included OSF](#) on list of acceptable data repositories
- Projects supported by FDA, USDA, NASA, and more bundle their outputs on OSF
- OSTP now requires each federal agency to develop guidance for funded research

Unique Persistent Identifiers

- ✔ Assigns PIDs to datasets
- ✔ PID points to persistent landing page

Long-Term Sustainability

- ✔ Long-term management of data
- ✔ Maintain availability of dataset
- ✔ Stable technical infrastructure
- ⚠ Stable funding
- ✔ Contingency plan for data

Metadata

- ⚠ Datasets must have metadata
- ⚠ Use schemas appropriate to the community

Free and Easy Access

- ✔ Free access to datasets and metadata
- ✔ Support for broad, equitable, open access
- ✔ Timely access after submission
- ✔ Maintain privacy, confidentiality, tribal sovereignty, and protection of sensitive data

Provenance

- ✔ Record the origin, chain of custody, and modifications

Curation and Quality Assurance

- ⚠ Datasets must have metadata

Broad and Measured Reuse

- ⚠ Measure attribution, citation, and reuse

Clear Use Guidance

- ✔ Clear documentation of terms for access and reuse

Security and Integrity

- ✔ Documented criteria for preventing unauthorized access, modification, or release of data
- ✔ Security levels appropriate to the sensitivity of data

Risk Management

- ✔ Ensure administrative, technical, and physical safeguards

Common Format

- ✔ Allows datasets and metadata downloaded, accessed, or exported
- ✔ Support for widely used and non-proprietary formats

Retention Policy

- ✔ Policy for data retention

Legend

- ✔ Characteristic met
- ⚠ Working towards characteristic

Metadata

Data shared without metadata isn't open

- Datacite metadata schema; Community-driven
- Open standards
- Controlled lists of resource types, funder, license, discipline, dates, authors, affiliations
- PID-ready (DOIs, ORCID iDs, ROR IDs)
- Relationships to connect outputs from across the research lifecycle
- New features:
 - New application profile and model based on international standards
 - Crossref funder lookup
 - File-level description and data types

The screenshot shows the OSF HOME interface for a project titled "Reproducibility Project: Cancer Biology". The page is categorized as a "Public project". A navigation sidebar on the left includes links for Overview, Metadata (which is highlighted), Files, Wiki, Resources, Components, Links, Analytics, Contributors, and Settings. The main content area is titled "Project Metadata" and contains several sections: "Description" with placeholder text, "Contributors" listing names like Timothy M. Errington and Sally N. Adebamowo, "Resource Information" showing the project type and language, "Funding/Support Information" with fields for funder, award title, and award number, and "Subjects".



a freely available generalist repository for all research disciplines and outputs

figshare.com

Flexibility

- Share any research output or file type
- Files up to 20GB, Support for big datasets
- Preview files in the browser
- Collections

Researcher Workflows

- Open API and FTP
- GitHub, GitLab, BitBucket Integrations
- Collaborative spaces
- Restricted Access

Persistent Metadata

- Unique DOI for each output, reservable
- ORCID integration
- Link to publications
- Link Funding via Dimensions

Open Access

- Open Access to all public files and metadata
- CC0 and CC-BY Licenses
- Discoverable across search engines, indexes
- FAIR commitment

Impact

- Public Author Profile
- Views, Downloads, Citations, Altmetrics
- Citations from full text literature
- Faceted Search



a FAIR repository for big data

plus.figshare.com

- Publish datasets over 20GB+ to 5TB or more
- File sizes up to 5TB
- Expert deposit support
- Dataset review
- One-time data publishing charge



Metadata

- DataCite Metadata Schema 4.4
- Unique DataCite DOIs for each item and collection
 - Version controlled
 - Reserve in advance
- Author ORCID
 - Multiple authors
 - Author profile
- Research Organization - *coming soon!*
- Funding sources: funder name, grant title and award number
- Fields of Research (FoR) 2020 subject categories
- Link related materials and publications with DOIs, URLs
- CC0, CC-BY Licenses, Software licenses, All metadata CC0
- Discoverable on Google, Google Dataset, Dimensions, DataCite, etc

Manage

needed to publish

Title
Vanderbilt Semantic Expertise Test Dataset

Authors
Ana Van Gulick × Isabel Gauthier × John Pyles ×
Teddy Van Gulick ×
Search co-authors by name, full email or ORCID. Hit enter after each.

Categories
Vision science and 2 more

Item type
Dataset

Keyword(s)
perceptual expertise × vision × individual differences ×
semantic knowledge × item analysis × MRI ×
behavioral measure ×
Add keywords for easy discovery. Hit enter after each

Description
Semantic Vanderbilt Expertise Test- SVET 1.0
A measure of semantic knowledge based on nomenclature, for the same categories as the VET 2.0.
As used in: Van Gulick, A. E., McGugin, R. W., & Gauthier, I. (2016). Measuring nonvisual knowledge about object categories: the semantic vanderbilt expertise test. Behavior research methods, 48(3), 1178-1196.
Five measures were used in Study 2, three domain-specific measures including an experience questionnaire (Gauthier et al., 2014), the SVET 1.0 as described in Study 1, and
H2 H3 H4 P B I U S |
≡ ≡ A₂ A¹

Funding
Do individual differences in face recognition predict perceptual expertise ×

+ Add another grant

References
http://gauthier.psy.vanderbilt.edu/resources/
https://doi.org/10.3758%2Fes13428-015-0637-5

Licence (what's this?)
CC BY 4.0

Tips
you can still drag more file(s) on the page or browse

Preview item (private)
Edit timeline

This item is a draft (metadata required for publication missing)

⌚ Apply embargo

🔗 Generate private link

🆔 Digital Object Identifier

10.6084/m9.figshare.20493363

The DOI becomes active when the item is published.

Licence (what's this?)

CC BY 4.0

CC BY 4.0

CC0

MIT

GPL

GPL 2.0+

GPL 3.0+

Apache 2.0



FAIR Open Data



Ana Van Gulick

[ORCID](#) [0000-0002-8133-6395](#)

Research Liaison for Psychology and Brain Sciences
(Psychology)
Carnegie Mellon University, Pittsburgh, PA

[Follow](#) [Twitter](#) [LinkedIn](#) [Facebook](#)

22845 item views **4300** item downloads **3** citations

Co-workers & collaborators

- John Borghi**
Data Librarian - Bay Area, CA
- Lisa Ziilinski**
Research Data Consultant - Pittsburgh, PA

11596
views

1682
downloads

87
citations [↗](#)

FUNDING

The Cognitive Neuroscience of Autism Spectrum Disorders

National Institute of Mental Health

[Find out more...](#)

Dimensions

DOCUMENTS [v](#) e.g. plastic AND instrument

Funding

R01EY02

Molecular Mechanisms of Human Retinal Disease
R01EY020540 | National Eye Institute

Genetic and Environmental Risk Factors for Exfoliation Syndrome and Glaucoma
R01EY020928 | National Eye Institute

Endoplasmic Reticulum Stress in Retinal Degeneration
R01EY020846 | National Eye Institute

Immunopathogenic mechanisms of dry eye disease
R01EY020889 | National Eye Institute

Formation and New Components of the Usher 2 Protein Complex in Photoreceptors
R01EY020853 | National Eye Institute

Gamma/delta T cells in autoimmune keratitis
R01EY021199 | National Eye Institute

Grant

The Cognitive Neuroscience of Autism Spectrum Disorders

Funder: National Institute of Mental Health (NIMH)
Grant number: ZIAMH002920 - [Original description](#)

Investigators

ALEX MARTIN - National Institute of Mental Health
PI

Research organization

National Institute of Mental Health, United States

Dataset DOI
Author Profiles
Author ORCID
Metrics
Associated Paper
Description for
discovery & reuse

Subject
License

Funding
Link to Awards



MENDELEY DATA supports **Research Data Management** workflow

Mendeley Data is a **free, secure, cloud-based generalist repository** where researchers can share data, ensuring it is easy to discover, access and cite.

Ensuring compliance and supporting open data, open science

Mendeley Data complies with FAIR data principles to ensure compliance to funder mandates. Metadata for all datasets published in Mendeley Data are available in an open format to facilitate large-scale acquisition and analysis of records—underscoring our commitment to open science.

Secure storage

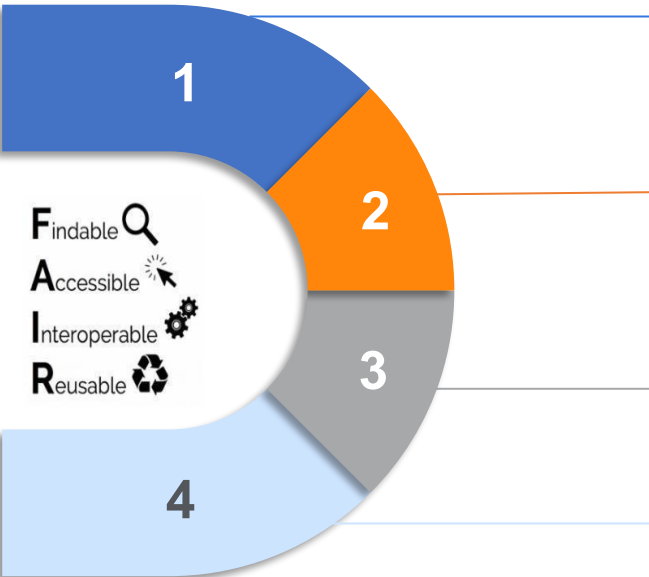
To ensure security, data is stored on Amazon's S3 servers. Our service was extensively penetration tested and received certification. Additionally, your valid published datasets are archived in perpetuity to preserve your data long-term. This ensures that the DOIs we provide for datasets will always resolve to a web page, where the dataset metadata and files will be available.

Easy curation

The Mendeley Data team supports the curation process before datasets are published. Datasets that do not meet the requirements for validity (scientific in nature and consisting of research data) will not become publicly visible. By reviewing your datasets prior to publication, and approving, editing, or returning datasets with comments, the Mendeley Data team helps to keep the quality of your content high.

Improving impact

Published dataset metadata is aggregated to DataCite's metadata index and to the OpenAIRE portal and indexed in Data Monitor. Metrics on downloads, views and data citations, sourced from DataCite and Crossref, are displayed on each dataset, as well as metrics capturing social media attention.



MENDELEY DATA for Data Sharing

<https://data.mendeley.com/>

Mendeley Data Homepage

Mendeley Data

Share your research data

Mendeley Data is a free and secure cloud-based communal repository where you can store your data, ensuring it is easy to share, access and cite, wherever you are.

Create a Dataset

Find out more about our institutional offering, [Digital Commons Data](#)

Search the repository

Find research data



Advanced search

Search results powered by [Data Monitor](#)

Recently published

Supplementary Material - Analysis of microbial diversity in ECZTRA 1, a randomized controlled trial

Lisa Beck, Thomas Bieber, Stephan Weidinger, Marie Tauber, Hidehisa Saeki et al
Published 29 November 2022 | Mendeley Data

Supplementary information to accompany and support the manuscript: Tralokinumab treatment improves the skin microbiota by increasing the microbial diversity in adults with moderate-to-severe atopic dermatitis: analysis of microbial diversity in ECZTRA 1, a randomized controlled trial.

Preview

Dataset Text

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



Create account

Sign in



The Generalist Repository Ecosystem Initiative

Elsevier's Mendeley Data repository is a participating member of the National Institutes of Health (NIH) Office of Data Science Strategy (ODSS) GREI project. The GREI includes seven established generalist repositories funded by the NIH to work together to establish consistent metadata, develop use cases for data sharing, train and educate researchers on FAIR data and the importance of data sharing, and more.

[Find out more](#)

Dataset Creation Interface

Draft (Version 3)

* Indicated fields which must be completed before publishing

Title *

Dataset on Mendeley Data

Contributors *

Luca Belletti × + Add contributor

Description *

text

2996 characters left

My files

New Folder Move Delete

- research-data ...
- research-data-1.xlsx 3 Oct 2022 16:57 9 KB ...
- research-data-2.xlsx 3 Oct 2022 16:57 9 KB ...

Drag files or folders here

Click here to upload

Published information

Status: Draft (Version 3)

Published version: Version 2

Visibility: Open access

Preview

Share

Delete

Publish

Reserved DOI: 10.17632/fr6s5x54mp.3

Cite this dataset

Belletti, Luca (2022), "Dataset on Mendeley Data", Mendeley Data, V3, doi: 10.17632/fr6s5x54mp.3

*DOI is reserved but not active

Copy to clipboard

Storage available

Space available: 10 GB

<https://www.elsevier.com/rdm>

<https://data.mendeley.com>

POLL

Anonymous



Best practice is to review repository guidelines prior to initiating a submission.
What is the most useful way for this information to be presented to make the process efficient? A submission checklist, instructional video, FAQs?



Breakout Session

30 min



Here's the plan:

1. Everyone will be randomly assigned a 'room' to join.
2. Facilitators will ask questions for your **feedback** and/or you can pose your own **questions** for discussion.
3. After **30 minutes**, we will regroup in the main meeting room.

Thank you for participating!
Have fun.



Breakout Session

30 min



Group 1: [Ana Van Gulick, Figshare](#)

Group 2: [Gretchen Gueguen, OSF](#)

Group 3: [Jess Herzog, Dryad](#)

Group 4: [Julie Wood, Vivli](#)

Group 5: [David Scherer, Mendeley](#)

Group 6: [Sara Gonzales, Zenodo](#)

Group 7: [Sonia Barbosa, Dataverse](#)



POLL

Anonymous



Based on what you've learned today, how do you feel about preparing data for a generalist repository?





Thanks for your
attention &
participation today!

Contact the GREI program
at GREI@nih.gov



Resources

Quickstart guide to data sharing

<https://datadryad.org/docs/QuickstartGuideToDataSharing.pdf>

GREI Collaborative Webinars (Sept-Dec 2022)

Presentation slides & recordings: <https://doi.org/10.17605/OSF.IO/JZU37>

