

Introduction to Using the Harvard Dataverse Repository & the NIH DMP Harvard Dataverse Guidance Document

Agenda

- The Harvard Dataverse Repository for Data Sharing and Use:
 - The Dataverse Project-Overview
 - The Dataverse Project and FAIR
 - Getting Started with The Harvard Dataverse Repository - Overview and Detailed
 - Collections
 - Datasets
 - Files/Sensitive Data
 - Search and Browse
 - Data Re-Use
 - NIH-GREI related improvements-forthcoming
- Overview of NIH DMP Guidance Document for Harvard Dataverse Repository
- Questions and Conversations

The Dataverse Project - Overview

The Dataverse Project

Open source research data repository software



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 norm for preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization tools, or other research and data archival systems with the Dataverse Project.*

The Dataverse Project

Community

- [Community Meetings](#)
- [Community Calls](#)
- [Global Dataverse Community Consortium](#)

Best Practices

- [Academic Credit](#)
- [Data Citation](#)
- [Dataverse Community Norms](#)
- [Data Management Guidelines](#)
- [Replication Dataset](#)

Software

- [Goals, Roadmap, and Releases](#)
- [Collaborations](#)
- [Integrations](#)
- [Features](#)
- [Source Code](#)
- [Guides](#)

DATVERSE REPOSITORIES - A WORLD VIEW



Dataverse for Data Management and Sharing

- Ease of data deposits
- Data citation
- Data access control
- Metadata standards
- Data sharing policies
- Data discovery

Supported Metadata

Detailed below are what metadata schemas we support for Citation and Domain Specific Metadata in the Dataverse Project:


- [Citation Metadata](#) (see .tsv version): compliant with DDI Lite, DDI 2.5 Codebook, DataCite 3.1, and Dublin Core's DCMI Metadata Terms . Language field uses ISO 639-1 controlled vocabulary.
- [Geospatial Metadata](#) (see .tsv version): compliant with DDI Lite, DDI 2.5 Codebook, DataCite, and Dublin Core. Country / Nation field uses ISO 3166-1 controlled vocabulary.
- [Social Science & Humanities Metadata](#) (see .tsv version): compliant with DDI Lite, DDI 2.5 Codebook, and Dublin Core.
- [Astronomy and Astrophysics Metadata](#) (see .tsv version): These metadata elements can be mapped/exported to the International Virtual Observatory Alliance's (IVOA) VOResource Schema format and is based on Virtual Observatory (VO) Discovery and Provenance Metadata (see .tsv version).
- [Life Sciences Metadata](#) (see .tsv version): based on ISA-Tab Specification, along with controlled vocabulary from subsets of the OBI Ontology and the NCBI Taxonomy for Organisms.
- [Journal Metadata](#) (see .tsv version): based on the Journal Archiving and Interchange Tag Set, version 1.2.

Experimental Metadata


Unlike supported metadata, experimental metadata is not enabled by default in a new Dataverse installation. Feedback via any [channel](#) is welcome!

- [CodeMeta Software Metadata](#): based on the CodeMeta Software Metadata Schema, version 2.0 (see .tsv version)
- [Computational Workflow Metadata](#) (see .tsv version): adapted from Bioschemas Computational Workflow Profile, version 1.0 and Codemeta.

Custom Metadata Blocks*


Roles  All the roles set up in your dataverse, that you can assign to users and groups.

[+ Add New Role](#)


Admin - A person who has all permissions for dataverses, datasets, and files, including approving requests for restricted data. 

[AddDataverse](#) [AddDataset](#) [ViewUnpublishedDataverse](#) [ViewUnpublishedDataset](#) [DownloadFile](#) [EditDataverse](#) [EditDataset](#) [ManageDataversePermissions](#) [ManageDatasetPermissions](#)

[ManageFilePermissions](#) [PublishDataverse](#) [PublishDataset](#) [DeleteDataverse](#) [DeleteDatasetDraft](#)


Contributor - For datasets, a person who can edit License + Terms, and then submit them for review. 

[ViewUnpublishedDataset](#) [DownloadFile](#) [EditDataset](#) [DeleteDatasetDraft](#)


Curator - For datasets, a person who can edit License + Terms, edit Permissions, and publish datasets. 

[AddDataverse](#) [AddDataset](#) [ViewUnpublishedDataverse](#) [ViewUnpublishedDataset](#) [DownloadFile](#) [EditDataset](#) [ManageDatasetPermissions](#) [ManageFilePermissions](#) [PublishDataset](#)


[DeleteDatasetDraft](#)

Dataset Creator - A person who can add datasets within a dataverse. 


[AddDataset](#)

Dataverse + Dataset Creator - A person who can add subdataverses and datasets within a dataverse. 


[AddDataverse](#) [AddDataset](#)

Dataverse Creator - A person who can add subdataverses within a dataverse. 

[AddDataverse](#)

File Downloader - A person who can download a published file. 

[DownloadFile](#)

Member - A person who can view both unpublished dataverses and datasets. 

[ViewUnpublishedDataverse](#) [ViewUnpublishedDataset](#) [DownloadFile](#)

Fe

Dataverse Repository Features

- Support for FAIR Data Principles
- Data citation for datasets and files
- OAI-PMH (Harvesting)
- APIs for interoperability and custom integrations
- API client libraries
- DataCite integration
- Login via Shibboleth
- Login via ORCID, Google, GitHub, or Microsoft
- Login via OpenID Connect (OIDC)
- Internationalization
- Versioning
- Restricted files
- Embargo
- Custom licenses
- Custom terms of use
- Publishing workflow support
- File hierarchy
- File previews
- Preview and analysis of tabular files
- Usage statistics and metrics
- Guestbook
- Fixity checks for files
- File download in R and TSV format
- Faceted search
- Customization of collections
- Private URL
- Widgets
- Notifications
- Schema.org JSON-LD
- External tools
- External vocabulary
- Dropbox integration
- GitHub integration
- Integration with Jupyter notebooks
- User management
- Curation status labels
- Branding
- Backend storage on S3 or Swift
- Direct upload and download for S3
- Export data in BagIt format
- Post-publication automation (workflows)
- Pull header metadata from Astronomy (FITS) files
- Provenance
- Support for rsync
- Auxiliary files for data files

<https://dataverse.org/software-features>

Data, Tools, Discoverability and Preservation

Getting Data Into a Dataverse Repository

- GitHub
- Dropbox
- OSF
- RSpace
- OJS
- Renku
- Amnesia
- SampleDB

Analysis and Computation

- Data Explorer
- Compute Button
- Whole Tale
- Binder
- Renku
- Avgidea Data Search

Discoverability

- OAI-PMH (Harvesting)
- SHARE
- Geodisy

Preservation

- Archivematica
- RDA BagIt Archiving
- Future Integrations*

The Dataverse Project and FAIR

FAIR Principles and Dataverse

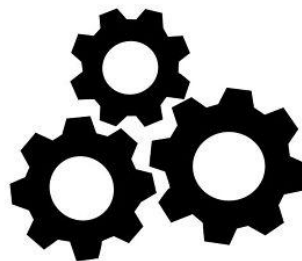
F_{indable}



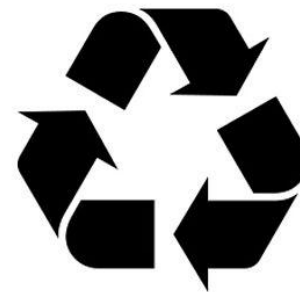
A_{ccessible}



I_{nteroperable}



R_{eusable}



https://en.wikipedia.org/wiki/File:FAIR_data_principles.jpg

How Dataverse Supports FAIR

- **FINDABLE:** Persistent Identifiers (DOIs, Handles) for datasets (and files), and enabling metadata indexing by search engines
- **ACCESSIBLE:** Open access to (metada)data and ensures data can be downloaded in machine-readable formats
- **INTEROPERABLE:** Provides standardised metadata schemas and enabling the integration of data with research tools and platforms
- **REUSABLE:** Licenses that clearly state how data can be used and ensuring that data are well documented and preserved for long term use

Getting started with the Harvard Dataverse Repository - Overview

The Purpose of a Data Repository

The **purpose of a data repository** is to keep a certain population of data so that it can be mined for greater insight or business intelligence or to be used for a specific reporting need.

https://en.wikipedia.org/wiki/Information_repository

The **Registry of Research Data Repositories (re3data.org)** is an **Open Science** tool that offers researchers, funding organizations, libraries and publishers an overview of existing international **repositories** for **research data**.

Data should be submitted to discipline-specific, community-recognized repositories where possible. **Where a suitable discipline-specific resource does not exist, data should be submitted to a generalist repository.**

 Metrics

52,530,800 Downloads

[Contact](#) [Share](#)

Share, archive, and get credit for your data. Find and cite data across all research fields.


[Advanced Search](#)
[+ Add Data](#)
☒ [Dataverses \(6,287\)](#)
☒ [Datasets \(166,066\)](#)
☐ [Files \(2,206,035\)](#)
Dataverse Category
[Research Project \(2,334\)](#)
[Researcher \(1,942\)](#)
[Organization or Institution \(503\)](#)
[Research Group \(441\)](#)
[Journal \(138\)](#)
[More...](#)
Metadata Source
[Harvard Dataverse \(93,652\)](#)
[Harvested \(78,701\)](#)
Publication Year
[2023 \(6,044\)](#)
[2022 \(19,842\)](#)
[2021 \(23,911\)](#)
[2020 \(9,906\)](#)

1 to 10 of 172,353 Results

[Sort](#)
Replication Data for: Just Patronage? Familiarity and the Diplomatic Value of Non-Career Ambassadors


Sep 13, 2023

 Michael Goldfien, 2023, "Replication Data for: Just Patronage? Familiarity and the Diplomatic Value of Non-Career Ambassadors", <https://doi.org/10.7910/DVN/MEOZ0E>, Harvard Dataverse, V2, UNF:6:LQnd+5Cx08z/SFzaeqtvHA== [fileUNF]

Career diplomats have expertise. Why, then, do U.S. presidents appoint relative novices to key diplomatic posts? Conventional wisdom points to patronage. Yet this explanation overlooks the benefits of a diplomat's familiarity with political superiors. Inherent in delegated diplom...

Replication Data for: Voters and the IMF: Experimental Evidence from European Crisis Countries

 Sep 13, 2023 - [Comparative Political Studies Dataverse](#)

 Hübscher, Evelynne; Sattler, Thomas; Wagner, Markus, 2023, "Replication Data for: Voters and the IMF: Experimental Evidence from European Crisis Countries", <https://doi.org/10.7910/DVN/6HMLST>, Harvard Dataverse, V1, UNF:6:WwffoWFahk/A52Mx2Giwwww== [fileUNF]

IMF interventions are often associated with rising political discontent in countries where the Fund intervenes. Studies examining this claim, however, face the challenge of disentangling the impact of the IMF from the impact of the crisis that triggered the intervention. To addre...

UK House of Commons Election Results at Constituency Level


Sep 13, 2023

 Umit, Resul, 2022, "UK House of Commons Election Results at Constituency Level", <https://doi.org/10.7910/DVN/S83HOA>, Harvard Dataverse, V7, UNF:6:y0vrHEGcToKy+I27B007ww== [fileUNF]

Visit the [Harvard Repository Policy page](https://support.dataverse.harvard.edu/policies) before getting started: <https://support.dataverse.harvard.edu/policies>

Prepare Your Data for Sharing in The Harvard Repository

GATHER AND VERIFY

Determine **what will be shared** and **ensure the metadata describes what is being shared**: raw data, output data, documentation, readme, codebook, etc...

Can you **share the data openly**? **Copyright?** **Deidentification?** (human subjects, vulnerable/endangered species); **Access restrictions** needed (restrict. embargo)

What **metadata standards** are required to describe your data properly?

***Sensitive data support is not yet available in Harvard Dataverse Repository. All data must be deidentified!**

CHOOSE AND ORGANIZE

Nonproprietary file formats for easy access, interoperability, preservation, plain text, same file but in open format version...

Organize the data by check all files for corruption, detailed file names, and usefulness. **Folders** should be clearly labeled and organized within zip files.

Know the **file size limitations in HDV** (Total project size varies 1TB to 2.5TB free, file size 2.5GB)

Tabular data get additional functionality in HDV

Know the **tools** available for use in HDV

SHARE YOUR DATA

Determine the repository (Domain specific should always be the first choice)

Share in one repository for one DOI (different components can be shared between domain specific and generalists but avoid sharing the same component in multiple repos)

Documentation such as README files should describe the data thoroughly and completely. Test your readme file instructions for reproducibility to ensure they are correct.

Share your citation and collection space URL

***all deposits must contain reusable data**

Create a Collection, Deposit Datasets, and Upload Files

Current Features



Dataverse Collections

- Own administration
- Own branding (and can be embedded anywhere)



Datasets

- Citation
- Metadata
- Versioning
- Private URL
- Custom Terms/Multiple License/Permissions
- Guestbooks
- Publishing Workflows



Files

- Citation
- Ingest
- Preview/Explore
- Metadata
- Versioning
- Permissions/Embargo/Restrictions

Log In

Log in or sign up with your institutional account — more information about account creation. Leaving your institution? Please contact Harvard Dataverse Support for assistance.

Your Institution



Harvard University

Please select...

Continue

Allow me to type the name of my institution

Other options

Username/Email

GitHub

Google

ORCID

Sign up for a Dataverse account.

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

Harvard Dataverse >

Contact Share Link Edit

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.

Search this dataverse... Advanced Search + Add Data

1 to 1 of 1 Result II Sort

Brain Genomics Superstruct Project (GSP)
May 15, 2015

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

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...

Publication Year
2014 (1)

Subject
Other (1)

Author Name
Buckner, Randy L. (1)
Hoffman, Joshua L. (1)
Smoller, Jordan W. (1)

Author Affiliation
Harvard University (1)



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

General Information

Theme + Widgets

Permissions

Groups

Dataset Templates

Dataset Guestbooks

Featured Dataverses

Current Features



Dataverse Collections

- Own administration
- Own branding (and can be embedded anywhere)



Datasets

- Citation
- Metadata
- Versioning
- Private URL
- Custom Terms/Multiple License/Permissions
- Guestbooks
- Publishing Workflows



Files

- Citation
- Ingest
- Preview/Explore
- Metadata
- Versioning
- Permissions/Embargo/Restrictions

Harvard Dataverse
Brain Genomics Superstruct Project (GSP) Dataset

Brain Genomics Superstruct Project (GSP) Dataset
(Harvard University)

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

Brain Genomics Superstruct Project (GSP)

Version 10.5

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.

Access Dataset +
Contact Owner Share

Dataset Metrics
13,076 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

Dataset Metrics

13,076 Downloads

Supported Metadata

Detailed below are what metadata schemas we support for Citation and Domain Specific Metadata in the Dataverse Project:

- **Citation Metadata** (see .tsv version): compliant with DDI Lite, DDI 2.5 Codebook, DataCite 3.1, and Dublin Core's DCMI Metadata Terms. Language field uses ISO 639-1 controlled vocabulary.
- **Geospatial Metadata** (see .tsv version): compliant with DDI Lite, DDI 2.5 Codebook, DataCite, and Dublin Core. Country / Nation field uses ISO 3166-1 controlled vocabulary.
- **Social Science & Humanities Metadata** (see .tsv version): compliant with DDI Lite, DDI 2.5 Codebook, and Dublin Core.
- **Astronomy and Astrophysics Metadata** (see .tsv version): These metadata elements can be mapped/exported to the International Virtual Observatory Alliance's (IVOA) VOResource Schema format and is based on Virtual Observatory (VO) Discovery and Provenance Metadata.
- **Life Sciences Metadata** (see .tsv version): based on ISA-Tab Specification, along with controlled vocabulary from subsets of the OBI Ontology and the NCBI Taxonomy for Organisms.
- **Journal Metadata** (see .tsv version): based on the Journal Archiving and Interchange Tag Set, version 1.2.

Subject ID Other
License/Data Use Agreement Custom Dataset Terms

Files Metadata Terms Versions

Search this dataset...

Filter by File Type All Access All File Tag All

1 to 10 of 10 Files

Download Request Access

GSP_Dataset_Terms_104022.pdf
Author: GSP, 2017-10-10
Published Aug 24, 2014
1,104 Downloads

GSP_101_140503.csv
File Type: CSV+XLS
Published Aug 24, 2014
1,104 Downloads

GSP_part10_140503.csv
File Type: CSV+XLS
Published May 21, 2014
1,104 Downloads

Cite Dataset + Learn about Data Citation Standards.

Author *
Barbosa, Sonia
Harvard University

Identifier Scheme *
Select...
ORCID
ISNI
LCNA
VIAF
GND
DAI
ResearcherID
ScopusID

Affiliation *
Harvard University

Contact *

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

Current Features



Dataverse Collections

- Own administration
- Own branding (and can be embedded anywhere)



Datasets

- Citation
- Metadata
- Versioning
- Private URL/*Anonymous Peer Review
- Custom Terms/*Multiple Licenses/Permissions
- Guestbooks
- Publishing Workflows



Files

- Citation
- Ingest
- Preview/Explore
- Metadata/Provenance
- Versioning
- Permissions/Embargo/Restrictions

The collage consists of five screenshots from the Dataverse interface:

- Top Left:** A screenshot of the 'Subject' page for a dataset, showing 'Files', 'Metadata', 'Terms', and 'Versions' tabs. It lists files like 'GSP_data10_140000.pdf' and 'GSP_data11_140000.pdf' with download links.
- Top Right:** A screenshot of the 'Files' page for a dataset, showing a list of files with details like 'AcqData.tab', 'alliance_v4.1_by_directed.tab', and 'alliance_v4.1_by_dyad.tab'. It includes download links and a 'File Access' dropdown menu.
- Middle Left:** A screenshot of the 'Edit Files' dropdown menu, showing options: 'Metadata', 'Restrict', 'Unrestrict', 'Tags', 'Embargo', and 'Delete'.
- Middle Right:** A screenshot of the 'Citation' page, showing the citation text for a specific dataset and a link to 'Learn about Data Citation Standards'.
- Bottom:** A screenshot of a data table with columns: 'staterno', 'stateab', 'year', 'total', 'land', 'sea', and 'version'. It displays data for various years and states, with a 'version' column showing values like 3.2.

Get Started on the Harvard Dataverse Repository - Detailed

Account Creation

➡ Log In

Log in or sign up with your institutional account — more [information about account creation](#). Leaving your institution? Please contact [Harvard Dataverse Support](#) for assistance.

Your Institution



HARVARD
UNIVERSITY

[Harvard University](#)

Please select... ▾

Continue

[Allow me to type the name of my institution](#)

Other options

Username/Email

GitHub

Google

ORCID

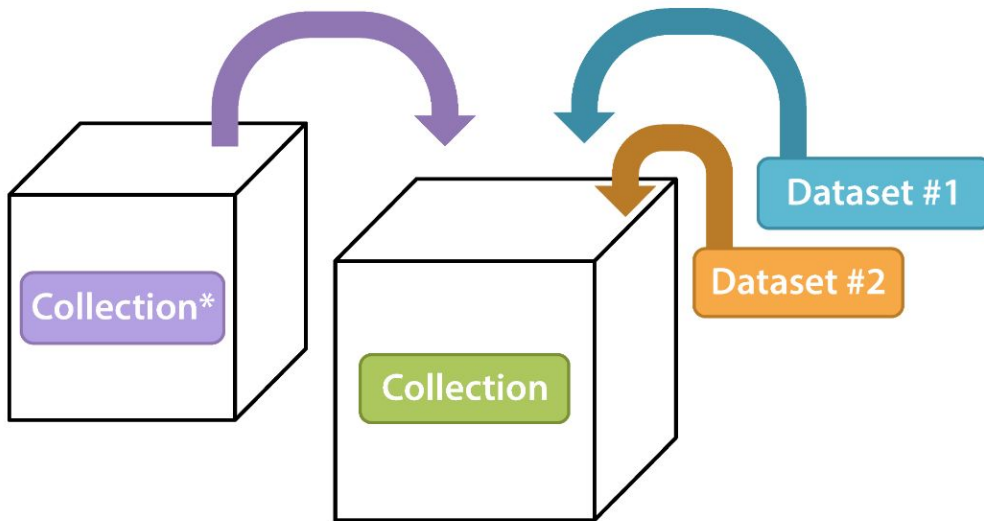
Create a Dataverse

“Collection”

- Edit Dataverse Collection
 - General Information
 - **Theme**
 - Widgets
 - Dataverse Collection Search Box Widget
 - Dataverse Collection Listing Widget
 - Adding Widgets to an OpenScholar Website
 - **Roles & Permissions**
 - Setting Access Configurations
 - Assigning Roles to Users and Groups
 - Dataset **Templates**
 - Dataset **Guestbooks**
 - Featured Dataverse Collection
- Dataset Linking
- Dataverse Collection Linking
- Publish Your Dataverse Collection

A Dataverse collection is a container for datasets (research data, code, documentation, and metadata) and other Dataverse collections, which can be setup for individual researchers, departments, journals and organizations.

Schematic Diagram of a **Collection** in Dataverse Software 5.0



Container for your **Datasets** and/or **Collections***

* Collections can contain other Collections



*contact support@dataverse.harvard.edu and request the creation of a collection

- Metadata
- Searchable facets
- Holds datasets



Customizable with descriptions, logos, collection space URLs, affiliation, contact emails, permissions, metadata and searchable facets. **Dataverse** collections hold datasets.

HARVARD Dataverse

Add Data ▾ Search ▾ About User Guide Support  Sonia Barbosa 

New Dataverse


*Asterisks indicate required fields

Host Dataverse ⓘ

Dataverse Name ⓘ

Identifier ⓘ

Category ⓘ

Email ⓘ
 

Affiliation ⓘ

Storage ⓘ

Description ⓘ
 This field supports only certain [HTML tags](#).

Metadata Fields

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

☒ Use metadata fields from Harvard Dataverse

☐ Citation Metadata (Required) [\[+\] View fields](#)

☐ 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](#)

Browse/Search Facets

Choose the metadata fields to use as facets for browsing datasets and dataverses in this dataverse.

☒ Use browse/search facets from Harvard Dataverse

All Metadata Fields ▾

		Selected
Keyword Term	→	Subject
Topic Classification Term	→	Author Name
Language	→	Author Affiliation
Producer Name	→	
Production Date	→	
Contributor Type	→	
Contributor Name	→	
Grant Information Grant	→	

Dataverse

Upgrade in progress.

Q

About

Support +

Contact

Dataverse Admin

Harvard Dataverse

A collaboration with Harvard Library, Harvard University IT, and IQSS

Harvard Dataverse

Permissions

Permissions

Here is the current access configuration to your dataverse.

Edit Access

Who can add to this dataverse?
Anyone adding to this dataverse needs to be given access

What should be the default role for someone adding datasets to this dataverse?
Contributor - Edit metadata, upload files, and edit files, edit Terms, Guestbook, Submit datasets for review

Users/Groups

Here are all the users and groups that have access to your dataverse.

Create Group

Assign Roles to Users/Groups

User/Group Name (Affiliation)	ID	Role	Action
Dataverse Admin (Dataverse.org)	@admin	Admin	<div>Remove Assigned Role</div>

Roles

Dataverse

Upgrade in progress.

Q

About

Support +

Contact

Dataverse Admin

Harvard Dataverse

A collaboration with Harvard Library, Harvard University IT, and IQSS

Permissions

Who can add to this dataverse?

☒ Anyone adding to this dataverse needs to be given access

☐ Anyone with a Dataverse account can add sub dataverses

☐ Anyone with a Dataverse account can add datasets

☐ Anyone with a Dataverse account can add sub dataverses and datasets

What should be the default role for someone adding datasets to this dataverse?

☒ Contributor - Edit metadata, upload files, and edit files, edit Terms, Guestbook, Submit datasets for review

☐ Curator - Edit metadata, upload files, and edit files, edit Terms, Guestbook, File Restrictions (Files Access + Use), Edit Permissions/Assign Roles + Publish

Save Changes

Cancel

Dataverse Admin (Dataverse.org)

@admin

Admin

Remove Assigned Role

Dataverse

Upgrade in progress.

Q

About

Support +

Contact

Dataverse Admin

Harvard Dataverse

A collaboration with Harvard Library, Harvard University IT, and IQSS

Assign Role

Grant permissions to users and groups by assigning them a role.

User/Group *

Test Testerson @testerson

Role *

☐ Admin

☐ Contributor

☐ Curator

☐ Dataset Creator

☐ Dataverse + Dataset Creator

☐ Dataverse Creator

☐ File Downloader

☐ Member

Save Changes

Cancel

Dataverse

Upgrade in progress.

Q

About

Support +

Contact

Dataverse Admin

Harvard Dataverse

A collaboration with Harvard Library, Harvard University IT, and IQSS

Assign Role

Grant permissions to users and groups by assigning them a role.

User/Group *

Test Testerson

Role *

☐ Admin

☒ Contributor

☐ Curator

☐ Dataset Creator

☐ Dataverse + Dataset Creator

☐ Dataverse Creator

☐ File Downloader

☐ Member

These are the permissions associated with the selected role.

Contributor

View/Upload/Delete Dataset

Download File

Edit Dataset

Delete Dataset Draft

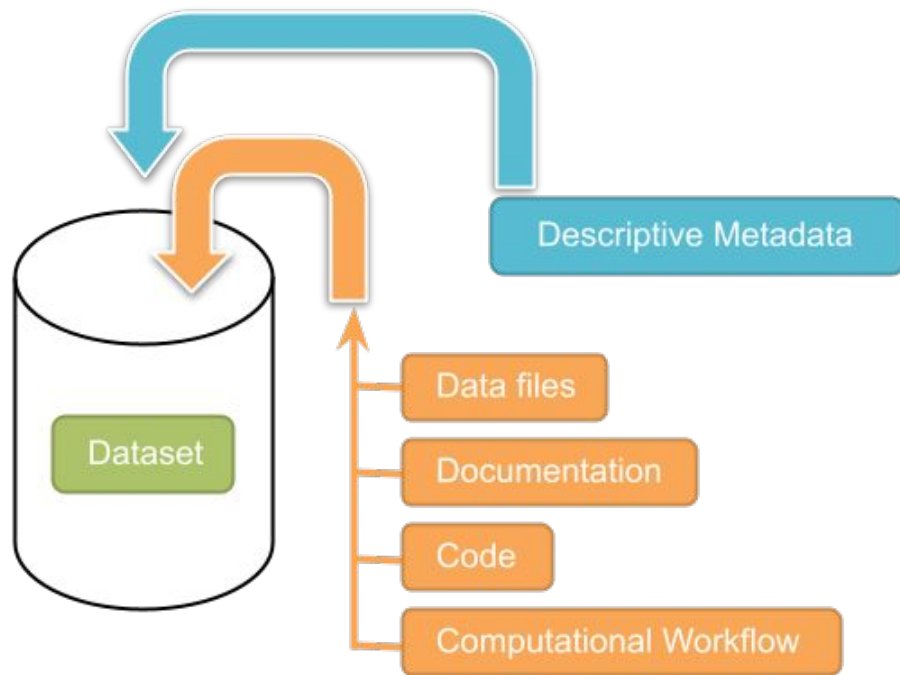
Save Changes

Cancel

Deposit a Dataset

- Supported Metadata
- Adding a New Dataset
- File Upload
- File Handling
- Restricted Files
- Edit Files
- Replace Files
- Terms
- Roles & Permissions
- Data Provenance
- Thumbnails + Widgets
- Publish Dataset
- Submit for Review
- Private URL to Review Unpublished Dataset
- Embargoes
- Dataset Versions
 - Version Details
- Dataset Metrics and Make Data Count
- Cloud Storage + Computing
 - Cloud Computing
 - Cloud Storage Access
- Dataset Deaccession

Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, code, and computational workflow.

Datasets - Metadata

Detailed below are what **metadata** schemas we support for Citation and Domain Specific Metadata in the Dataverse Project:

- **Citation Metadata**: compliant with [DDI Lite](#), [DDI 2.5 Codebook](#), [DataCite 3.1](#), and Dublin Core's [DCMI Metadata Terms](#) (see [.tsv version](#)). Language field uses [ISO 639-1](#) controlled vocabulary.
- **Geospatial Metadata**: compliant with DDI Lite, DDI 2.5 Codebook, DataCite, and Dublin Core (see [.tsv version](#)). Country / Nation field uses [ISO 3166-1](#) controlled vocabulary.
- **Social Science & Humanities Metadata**: compliant with DDI Lite, DDI 2.5 Codebook, and Dublin Core (see [.tsv version](#)).
- **Astronomy and Astrophysics Metadata**: These **metadata** elements can be mapped/exported to the International Virtual Observatory Alliance's (IVOA) [VOResource Schema format](#) and is based on [Virtual Observatory \(VO\) Discovery and Provenance Metadata](#) (see [.tsv version](#)).
- **Life Sciences Metadata**: based on [ISA-Tab Specification](#), along with controlled vocabulary from subsets of the [OBI Ontology](#) and the [NCBI Taxonomy for Organisms](#) (see [.tsv version](#)).
- **Journal Metadata**: based on the [Journal Archiving and Interchange Tag Set, version 1.2](#) (see [.tsv version](#)).

See also the [Dataverse Software 4.0 Metadata Crosswalk: DDI, DataCite, DC, DCTerms, VO, ISA-Tab](#) document and the [Metadata Customization](#) section of the Admin Guide.

Datasets - Metadata

Metadata Fields

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

☒ Use metadata fields from Harvard Dataverse

☒ Citation Metadata (Required) [\[+\] View fields](#)

☐ 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](#)

FilesMetadataTermsVersions

Export Metadata

Citation Metadata

Dataset Persistent ID

doi:10.7910/DVN/HADLCK

Publication Date

2021-05-26

Title

Assessment of Land Degradation in Semi-Arid Zone of Central Tanzania

Author

International Institute of Tropical Agriculture (IITA)
University of Bonn, Germany

Contact

Use email button above to contact.

Description

A sub-national field assessment of land degradation was conducted in the Kongwa districts of Tanzania in December 2019. 34 sampling plots were selected using a stratified sampling method based on a land cover map. One site that hosts Africa RISING technologies on land rehabilitation was purposely selected to act as a control. The primary sampling plots measured 100*100m and were subdivided into replicate sub-plots measuring 30x30m. A sub-sample of 3 sub-plots was selected in each primary plot for assessment of land degradation. The observations from 3 subplots (30*30) were averaged to obtain an aggregate value for the larger plot (100x100m). The antecedent biophysical conditions in the sampling plot were recorded i.e. the land use, degree of slope, topographical position, soil color, crop types grown and land tenure. The percentage of area that was undegraded in each plot was also estimated visually.

A questionnaire for mapping land degradation and sustainable land management was applied for visual assessment of the type, extent, degree, and direct causes of land degradation. The different types of land degradation, for example erosion by water, were scored whether they are present or not, in addition to their extent and degree (intensity). The extent represented the proportion of a sub-plot covered by different types of land degradation, the degree was divided into 4 classes in ascending order of intensity of land degradation (0 = Light, 1 = Moderate, 2 = Strong, 3 = Severe). Moreover, the type, purpose, extent, and effectiveness of sustainable land management practices (SLM) were visually assessed in every subplot. Data were recorded using the mobile-based KoboCollect toolbox and transmitted to a cloud database for storage and descriptive analysis.

Subject

Agricultural Sciences

Keyword

land degradation (AGROVOC) http://aims.fao.org/aos/agrovoc/c_34823
sustainable agriculture (AGROVOC) http://aims.fao.org/aos/agrovoc/c_33561
soil erosion (AGROVOC) http://aims.fao.org/aos/agrovoc/c_2651
TANZANIA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_7608
EAST AFRICA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_2442
AFRICA SOUTH OF SAHARA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_166
AFRICA (AGROVOC) http://aims.fao.org/aos/agrovoc/c_165

Topic Classification

Natural resources (AGROVOC) http://aims.fao.org/aos/agrovoc/c_5091
Sustainable agriculture (AGROVOC) http://aims.fao.org/aos/agrovoc/c_33561


Notes

If you are interested in accessing restricted datasets, kindly fill out this [data use agreement form](#).


Language

English


Producer

International Institute of Tropical Agriculture (IITA) <https://www.iita.org> 

Datasets - Citations

 HARVARD
Dataverse


Add Data ▾ Search ▾ About User Guide Support Sign Up Log In

 Africa RISING
Africa RISING Dataverse (International Food Policy Research Institute (IFPRI)) Africa RISING

[Harvard Dataverse](#) > [IFPRI Dataverse](#) > [Africa RISING Dataverse](#) >

Assessment of Land Degradation in Semi-Arid Zone of Central Tanzania

Version 1.0



International Institute of Tropical Agriculture (IITA); University of Bonn, Germany, 2021, "Assessment of Land Degradation in Semi-Arid Zone of Central Tanzania", <https://doi.org/10.7910/DVN/HADLCK>, Harvard Dataverse, V1, UNF:6:jU30uW/R5JgiHw7dPr5iuQ== [fileUNF]

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

Contact Owner Share


Dataset Metrics ⓘ
0 Downloads ⓘ

Brain Genomics Superstruct Project (GSP) Dataverse Home
(Harvard University)

[Harvard Dataverse](#) > [Brain Genomics Superstruct Project \(GSP\) Dataverse](#) >

Brain Genomics Superstruct Project (GSP)

Version 10.5



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.](#)

Access Dataset ▾
Edit Dataset ▾
Link Dataset
Contact Owner Share

Dataset Metrics ⓘ
13,981 Downloads ⓘ

Data - Citations

Your Publication



**Formal
Data
Citation**



Your Data

A screenshot of a data table with a grid of numerical values. The table has multiple columns and rows, with some cells containing text like '1.000' and '0.000'. The table is displayed in a window titled 'Data Table'.

Principle 2
- **Credit and Attribution:**

Such as authors, repositories or other distributors and contributors.

← **Author(s)**, Year, Dataset Title,
Global Persistent Identifier, →
← **Data Repository or Archive**,
Version



Principle 7 - Specificity and verification:
Such as the specific version used. Versioning or timeslice information should be supplied with any updated or dynamic dataset.

Principles 4, 5, 6
- **Unique Identification, Access, Persistence:**

A unique, persistent identifier, such as a DOI or Handle, that provides access to metadata.

Fig. 1 Example of a data citation based on the [Joint Declaration of Data Citation Principles \(2014\)](#).

Data - Citations

Related Material ?

Examples of prior publications of GSP data with partial data description:

Yeo, B.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., Fischl, B., Liu, H., Buckner, R.L. (2011) The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3): 1125-1165: [Link to article](#)

Buckner, R.L., Krienen, F.M., Castellanos, A., Diaz, J.C., Ye o, B.T. (2011) The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(5): 2322-2345: [Link to article](#)

Choi, E.Y., Yeo, B.T.T., Buckner, R.L. (2012) The organization of the human striatum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 108(8): 2242-2263: [Link text](#)

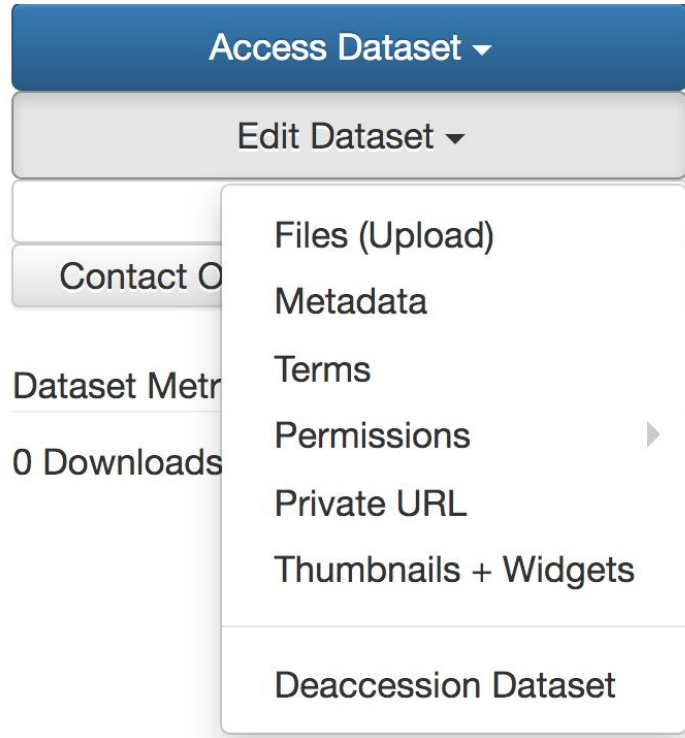
Van Dijk, K.R., Sabuncu, M.R., Buckner, R.L. (2012) The influence of head motion on intrinsic connectivity MRI. *NeuroImage* , 59(1): 431-438: [Link to article](#)

Holmes, A.J., Lee, P.H., Hollinshead, M., Bakst, L., Roffman, J.L., Smoller, J.W., Buckner, R.L. (2012) Individual differences in amygdala-prefrontal anatomy link negative affect, impaired social functioning, and polygenic depression risk. *Journal of Neuroscience*, 32(50): 18087-18100: [Link to article](#)

Related Publication ?

Furnas, Alexander, and Timothy LaPira. [date]. "The People Think What I Think: False Consensus and Unelected Elite Misperception of Public Opinion." *American Journal of Political Science* Forthcoming. <http://ajps.org/>

Datasets - Editing Options



- Datasets can be edited with the options seen here, and versioning is available for dataset and files.
- A **Private URL** can be generated for DRAFT datasets to share a dataset with a coauthor or journal, before publishing.
- **Deaccession** is the only option to “remove” a dataset once it has been published and this leave the TOMBSTONE page of the citation.

Datasets: Permissions, Workflows and Roles

Assign Role

Grant permissions to users and groups by assigning them a role.

Users/Groups *

Role *

- ☐ Admin
- ☐ Contributor
- ☐ Curator
- ☐ File Downloader
- ☐ Member

These are the permissions associated with the selected role.

Save Changes

Cancel

Grant File Access

Grant file access to users and groups.

Users/Groups

Selecting 0 of 20 Restricted Files

	File Name
<input type="checkbox"/>	00097_Early_Head_Start_B1C_ruf.sd2
<input type="checkbox"/>	00097_Early_Head_Start_B1C_ruf.tab
<input type="checkbox"/>	00097_Early_Head_Start_B1H_ruf.sd2
<input type="checkbox"/>	00097_Early_Head_Start_B1H_ruf.tab
<input type="checkbox"/>	00097_Early_Head_Start_B2C_ruf.sd2
<input type="checkbox"/>	00097_Early_Head_Start_B2C_ruf.tab
<input type="checkbox"/>	00097_Early_Head_Start_B2H_ruf.sd2
<input type="checkbox"/>	00097_Early_Head_Start_B2H_ruf.tab
<input type="checkbox"/>	00097_Early_Head_Start_B2K_ruf.sd2
<input type="checkbox"/>	00097_Early_Head_Start_B2K_ruf.tab

✓ Grant

Cancel

Datasets

Description ⓘ

A sub-national field assessment of land degradation was conducted in the Kongwa districts of Tanzania in December 2019. 34 sampling plots were selected using a stratified sampling method based on a land cover map. One site that hosts Africa RISING technologies on land rehabilitation was purposely selected to act as a control. The primary sampling plots measured 100*100m and were subdivided into replicate sub-plots measuring 30x30m. A sub-sample of 3 sub-plots was selected in each primary plot for assessment of land degradation. The observations from 3 subplots (30*30) were averaged to obtain an aggregate value for the larger plot (100x100m). The antecedent biophysical conditions in the sampling plot were recorded i.e. the land use, degree of slope, topographical position, soil color, crop types grown and land tenure. The percentage of area that was undegraded in each plot was also estimated visually.

[Read full Description \[+\]](#)

Subject ⓘ

Agricultural Sciences

Keyword ⓘ

land degradation, sustainable agriculture, soil erosion, TANZANIA, EAST AFRICA, AFRICA SOUTH OF SAHARA, AFRICA

Notes ⓘ

If you are interested in accessing restricted datasets, kindly fill out this [data use agreement form](#).

Files

Metadata

Terms

Versions

Q

Filter by

File Type: All ▾

Access: All ▾


File Tag: All ▾


↕ Sort ▾


☐

1 to 2 of 2 Files

☐



[Field_protocol](#)
Adobe PDF - 334.1 KB
Published May 26, 2021
0 Downloads
MD5: 041...e2d 



Datasets - Terms

[Files](#) [Metadata](#) [Terms](#) [Versions](#)

Terms of Use

Weaver

Terms of Use

IPPRI Datasets Terms of Use Statement

(December 2019)
1. Description

The International Food Policy Research Institute (IFPRI) views the products of its research including research data, global public goods and knowledge as essential to enabling their widespread distribution and use by providing unrestricted access. This is in keeping with the [IFPRI Research Data Management and Open Access \(ROMDAO\) Policy](#) and the [Access and Data Management Policy](#) (each, a "Policy" and together, the "Policies").

IFPRI makes all its research data publicly available through the "IFPRI Database". The IFPRI Database houses a comprised of datasets collected during the course of IFPRI research, including data tables in various standard file formats, codebooks, metadata, and other associated documentation. IFPRI Database is a network member database, which is a repository for research data hosted at [Institute for Quantitative Social Science \(IQSS\)](#), Harvard University.

IFPRI is the owner of the "IFPRI Database" (sometimes referred to in these Terms of Use as the "Site"). The Center for Global Public Affairs (CPA) division of IFPRI operates this Site on behalf of IFPRI. For questions pertaining to these licenses, or concerns, please contact IFPRI-Dataset@cgiar.org.

In the event of any conflict between these Terms of Use and either of the Policies, the terms of the applicable Policy shall prevail over and to the extent of, any such conflicting terms.

2. Adoption, Acceptance, and Changes to Terms of Use

By using this Site, downloading or otherwise accessing any dataset or other materials or content made available ("you" or "user"), acknowledge that you understand and intend these Terms of Use to be the legal equivalent of a written contract and equally binding, and that you accept the terms of and agree to be legally bound by these Terms of Use. IFPRI may change these Terms of Use without prior notification. However, any modifications to the Terms of Use effective for the download/use of this Site subsequent to such modification. Your continued download of dataset files from this Site subsequent to reasonable notice of such modification will be conclusively deemed acceptance of such modification to these Terms of Use. You agree that notice of changes to these Terms of Use posted on this Site is reasonable and sufficient notice. We consider it incumbent upon users to read these terms to be aware of changes. You are bound by the then-current version of these Terms of Use and all applicable laws. Accordingly, we recommend that you review these Terms of Use from time to time to ensure that you are familiar with the most recent version.

3. Use of the Materials/Site

By using this Site or downloading, or otherwise accessing the IFPRI Database, users acknowledge:

1. The dataset or any other materials made available through the site will be used solely for reporting aggregate information, and not for investigation of specific individuals or organizations.
2. IFPRI will not be responsible or held accountable for any interpretations or conclusions that users may draw from use of dataset or materials available on the Site.
3. No attempt will be made to identify research subjects from the data or materials that are accessed from the site, and no user will be made of the identity of any person, institutions or establishment discovered inadvertently. If user becomes aware of any violation of the foregoing, user will immediately report it to [IFPRI-Dataset](#).
4. No attempt will be made to remove links among datasets, datasets or any other materials that are available through this

Save Changes

Terms

Dataset Terms

License/Data

Terms of Use

Confidentiality

Special Permission

[illegible]

Dataset Terms

Licence/Data Use Agreement

This dataset will be published under the terms specified below. You must expect that proper credit is given via citation.

CC0 1.0

CC0 1.0

CC BY 4.0

CC BY-NC 4.0

CC BY-NC-ND 4.0

CC BY-NC-SA 4.0

CC BY-ND 4.0

CC BY-SA 4.0

PDDL-1.0

Restricted Files + Terms of Access

Info – Restricting limits access to publishing information about access to the Terms of Access

Multiple License support

Soybean Leaf Defoliation Data Set For DefoNet

Version 2.0

Zhang, Zichen, 2022, "Soybean Leaf Defoliation Data Set For DefoNet", <https://doi.org/10.7910/DVN/VJMBUO>, Harvard Database, V2

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

Description

Soybean leaf defoliation data set for a CNN model, DefoNet, used in my paper "Assessing the Efficacy of Machine Learning Techniques to Characterize Soybean Defoliation from Unmanned Aerial Vehicles". (2022-03-29)


Subject

Computer and Information Science; Agricultural Sciences

Related Publication

Zhang, Zichen, Sami Khanal, Amy Raudenbush, Kelley Tilmon, and Christopher Stewart. "Assessing the efficacy of machine learning techniques to characterize soybean defoliation from unmanned aerial vehicles." Computers and Electronics in Agriculture 193 (2022): 106682.

License/Data Use Agreement

 CC0 1.0







Files Metadata Terms Versions

Datasets - Versions

Files	Metadata	Terms	Versions
Dataset	Summary	Contributors	Published
1.0	This is the first published version.	IFPRI KM	May 26, 2021

<https://dataverse.org/files/dataverseorg/files/metadataprovenance-mercecosas.pdf>





Datasets - Files

<input type="checkbox"/>	 CenPop2010_Mean_TR.txt raw data/Census/ Plain Text - 3.0 MB Published Sep 13, 2023 0 Downloads MD5: 89a...1e2 	<p>Open access files will automatically utilize tools available for data discoverability</p>  
<input type="checkbox"/>	 accident2000.tab raw data/FARS/ Tabular Data - 5.5 MB Published Sep 13, 2023 0 Downloads 51 Variables, 37526 Observations UNF:6:GcWk...hqQ== 	

Files Metadata Terms Versions

☐

1 File

<input type="checkbox"/>	 CGIP_catla_silver_GxE.xlsx MS Excel Spreadsheet - 708.8 KB Published Sep 12, 2023 0 Downloads MD5: 8c5...59d 	<p>Restricted access workflows can allow for “request access” to data directly on the file page, or you can utilize other request access workflows under the TERMS tab</p> 
	<div>6. Raw and Verified Data</div>	

Datasets - Tabular Files


Supported File Formats


Tabular Data ingest supports the following file formats:

File format	Versions supported
SPSS (POR and SAV formats)	7 to 22
STATA	4 to 15
R	up to 3
Excel	XLSX only (XLS is NOT supported)
CSV (comma-separated values)	(limited support)

See the subsections in the left sidebar for more information on each of these supported formats.



Datasets - Tabular Files








[accident2000.tab](#)
raw data/FARS/
Tabular Data - 5.5 MB
Published Sep 13, 2023
0 Downloads
51 Variables, 37526 Observations UNF:6:GcWk...hqQ== 

Ask the Data










[accident2000.tab](#)
raw data/FARS/
Tabular Data - 5.5 MB
Published Sep 13, 2023
0 Downloads
51 Variables, 37526 Observations UNF:6:GcWk...hqQ== 

View Data







[accident2000.tab](#)
raw data/FARS/
Tabular Data - 5.5 MB
Published Sep 13, 2023
0 Downloads
51 Variables, 37526 Observations UNF:6:GcWk...hqQ== 





[accident2001.tab](#)
raw data/FARS/
Tabular Data - 5.6 MB
Published Sep 13, 2023
0 Downloads
51 Variables, 37862 Observations UNF:6:wai8...mog== 





[accident2002.tab](#)
raw data/FARS/
Tabular Data - 5.8 MB
Published Sep 13, 2023
0 Downloads
51 Variables, 38491 Observations UNF:6:9tkh...uQ== 





[accident2003.tab](#)
raw data/FARS/



File Access 

Public

Download Options 

Comma Separated Values (Original File Format)

Tab-Delimited

RData

Download Metadata 

Variable Metadata

Data File Citation

Explore Options 

Data Explorer

<https://guides.dataverse.org/en/latest/user/tabulardataingest/index.html>

Datasets - Ask the Data

Tell me what you want to know

What is this data about?

Answer please

This data is about traffic accidents in the US.

STATE	COUNTY	MONTH	DAY	HOUR	MINUTE	VE_FORMS	PERSONS	PEDS	NHS	ROAD_FNC	ROUTE	SP_JUR	HARM_EV	MAN_COLL	REL_JUNC
1	97	1	26	15	30	2	2	0	0	6	4	0	12	2	1
1	101	2	1	23	55	1	1	0	1	2	2	0	42	0	1
1	81	2	2	5	10	1	1	0	0	6	4	0	42	0	1
1	101	1	18	22	14	1	2	1	1	12	2	0	8	0	1
1	101	1	20	13	5	1	1	0	1	11	1	0	34	0	1
1	73	1	31	2	40	2	3	0	1	11	1	0	25	0	1
1	73	1	14	19	0	1	2	0	1	11	1	0	34	0	1
1	73	1	24	19	45	2	3	1	0	6	6	0	8	0	2
1	97	1	23	19	34	2	3	0	0	13	2	0	12	5	2
1	97	1	31	23	33	1	2	0	0	4	3	0	30	0	1

Viewing rows 1 through 10 of 38444

Datasets - File Previewer

PreviewMetadataVersions

Explore on View Data

	householdid	memname	age	reltohhh	otherreltohh	gender	educlvl	othereduclvl	priactvty16	othpriactvty16	priactvty18
5	2060154		25	3		1	/		14		14
6	2060154		18	3		2	5		14		14
7	2060156		61	1		1	4		6		1
8	2060156		17	3		2	6		14		13
9	2060156		15	3		1	6		14		16
10	2060156		32	2		2	8		18		18
11	2060157		52	1		1	4		5		5
12	2060157		37	2		2	4		5		5
13	2060157		32	3		1	6		5		5
14	2060157		32	3		2	4		8		8
15	2060158		48	1		1	6		5		5
16	2060158		25	2		2	4		8		8
17	2060158		17	3		1	5		14		14
18	2060158		16	3		1	5		5		14
19	2060159		48	1		2	4		7		5
20	2060159		21	3		1	5		6		14
21	2060159		17	3		2	5		6		14
22	2060159		15	3		2	4		2		13
23	2060160		80	1		1	3		3		4
24	2060160		50	2		2	4		6		4
25	2060160		30	3		1	5		8		16
26	2060160		25	3		2	4		5		16
27	2060161		70	1		1	4		8		8
28	2060161		50	2		2	1		8		5
29	2060161		30	3		1	6		5		8
30	2060152		27	3		1	6		5		14
31	2060153		48	1		2	4		1		8

Datasets - GeoJSON Previewer

This file is part of "GeoJSON Example".

Version 1.0

File Citation

Durbin, Philip, 2022, "wikipedia.geojson", GeoJSON Example, <https://doi.org/10.70122/FK2/GFBLSO/S7JASJ>, Demo Dataverse, V1

[Cite Data File](#)

[Learn about Data Citation Standards.](#)

Dataset Citation

Durbin, Philip, 2022, "GeoJSON Example", <https://doi.org/10.70122/FK2/GFBLSO>, Demo Dataverse, V1

[Cite Dataset](#)

[Learn about Data Citation Standards.](#)

Access File

Contact Owner

Share

File Metrics

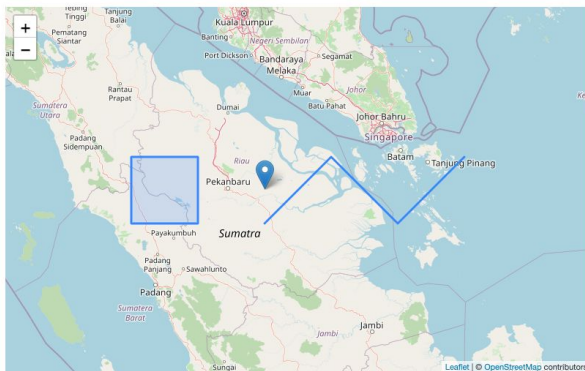
4 Downloads

Preview

Metadata

Versions

[Open in New Window](#)



Previewers originally developed by ODRI and maintained at <https://github.com/GlobalDataverseCommunityConsortium/dataverse-previewers>. Feedback and contributions welcome.

<https://github.com/GlobalDataverseCommunityConsortium/dataverse-previewers>

Datasets - Tabular Files/Data Explorer

Third National Fadama Development Financing II Impact Study Household Survey in Bauchi

004.SectionA1_HHD15yrsandabove.tab

International Food Policy Research Institute (IFPRI), 2021, "Third National Fadama Development Financing II Impact Study Household Survey in Bauchi", <https://doi.org/10.7910/DVN/AEROHZ>, Harvard Dataverse, V1, UNF:6:tu65LD7Bc80cPPs7DKCBHQ== [fileUNF]

Q	13 Results	Download	Chart View	Table View
ID	Name	Label	C	
22529304	householdid	household ID		
22529305	memname	Name of HH member who is above 15 years old		
22529302	age	Age of member		
22529309	reltohh	Relationship to household head	7	
22529307	otherreltohh	Other relationship to household head (specify)		
22529311	gender	Gender of member	2	
22529301	educvl	Highest Level of education	9	
22529299	othereducvl	Other level of education(specify)		
22529303	practvty16	Primary activity in 2016	2	
22529308	othpractvty16	Other primary activity(specify)in 2016		

Variable educvl: Highest Level of education

Values	Categories
8	Koranic education
5	Some secondary education (incl. Junior secondary school)
2	Adult literacy training
1	No formal education
99	Others
7	Post-secondary education
6	Completed secondary education
3	Some primary education
4	Completed primary education

Summary Statistics

Variable othpractvty16: Other primary activity(specify)in 2016

Values	Categories
--------	------------

Summary Statistics

Variable practvty16: Primary activity in 2016

Values	Categories
14	Student in school (any type)
2	Livestock production
7	Transportation business
10	Construction
18	Artisans(Incl. Mechanics)
3	Fisheries
12	Public sector employment
13	Domestic duties
4	Forest production and/or harvesting
15	Retired

<https://guides.dataverse.org/en/latest/user/tabulardataingest/index.html>
<https://github.com/scholarsportal/dataverse-data-explorer-v2>


<https://guides.dataverse.org/en/latest/user/tabulardataingest/index.html>


<https://github.com/scholarsportal/dataverse-data-explorer-v2>

Datasets - File Level Embargo


Access: EmbargoedThenPublic ✕


1 to 4 of 4 Results Sort ▾

 Kenya_COVID19_R4_codebook_responsecode.tab Embargoed


Mar 24, 2022 - COVID-19 Impact on Rural Men and Women in Kenya, Round 4
Tabular Data - 22.3 KB - 6 Variables, 648 Observations - UNF:6:9Mrp...qEW== 


Codebook Documentation

 Kenya_COVID19_R4_codebook.tab Embargoed


Mar 24, 2022 - COVID-19 Impact on Rural Men and Women in Kenya, Round 4
Tabular Data - 25.3 KB - 7 Variables, 285 Observations - UNF:6:bV12...CVQ== 


Codebook Documentation

 Kenya_COVID19_R4_questionnaire.tab Embargoed


Mar 24, 2022 - COVID-19 Impact on Rural Men and Women in Kenya, Round 4
Tabular Data - 51.6 KB - 23 Variables, 333 Observations - UNF:6:n+Ae...L3A== 

Documentation Questionnaire

 Kenya_COVID19_R4.tab Embargoed

Mar 24, 2022 - COVID-19 Impact on Rural Men and Women in Kenya, Round 4
Tabular Data - 243.3 KB - 285 Variables, 507 Observations - UNF:6:dZz1...M6Q== 

Data Panel Survey

Edit Options 

Metadata

Restrict

Replace

Embargo

Delete


Harvard Dataverse > IFPRI Dataverse >

COVID-19 Impact on Rural Men and Women in Kenya, Round 4

Embargoed Version 1.0

File Metadata ▴

Preview



File Tags

File UNF

Original File MD5

Deposit Date

Metadata Release Date

Publication Date

Size

Type

Variables

Observations

Data Panel Survey

UNF:6:dZz15Xn5h4XY+gy4SyOM6Q==

d892505e28783d6b5be3cf1588f9e148

2022-03-01

2022-03-24

Embargoed until 2023-07-31

243.3 KB

Tab-Delimited

285

507

Datasets - Deaccession

Replication Data for: Democratic Subversion: Elite Cooptation and Opposition Fragmentation

Deaccessioned



Jun 8, 2021

Arriola, Leonardo R.; DeVaro, Jed; Meng, Anne, 2021, "Replication Data for: Democratic Subversion: Elite Cooptation and Opposition Fragmentation", <https://doi.org/10.7910/DVN/GHDKSY>, Harvard Dataverse, V1, DEACCESSIONED VERSION, UNF:6:Nk/xhGPVreurGI55KFcKrQ== [fileUNF]

The dataset has been transferred to another repository

Considerations in Sharing Sensitive Data in Harvard Dataverse Repository

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

Terms of use and application form for data requests

***Sensitive data support is not yet available in Harvard Dataverse Repository.**
All data must be deidentified!

Children of Gay Fathers, 1985

Version 3.2

Bozett, Frederick W., 2022, "Children of Gay Fathers, 1985", <https://doi.org/10.7910/DVN/D24VWQ>, 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 fathers' homosexuality, how they feel about it, who and why, whom they haven't told and why, what other people's reactions have been, and how their relationship with their father is any different since they found out that he is gay. What acts in any way 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...



Access Dataset

[Contact Owner](#)

[Share](#)

Dataset Metrics

112 Downloads



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: f67...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: f6e...94c

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...b0f

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://thedata.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.

Utilize the “guestbook” feature

Restricted, de-identified data

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

Search and Browse Functionality

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)

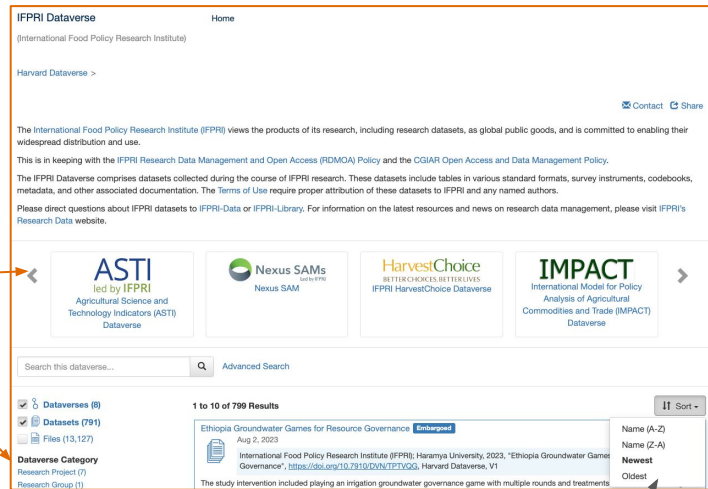
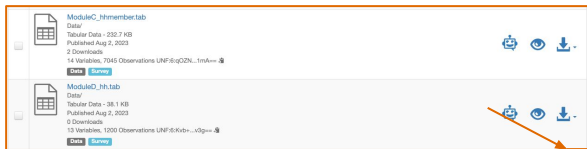


Dataset details:

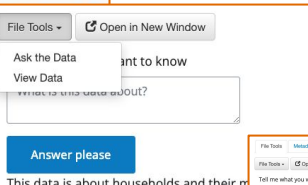
- Detailed dataset page
- Metadata
- Description
- Authors
- Citation
- Download options
- Documentation, readme, code...
- Multiple format download options

Browse:

- Categories/Subjects
- Featured Datasets/collections



Ask the Data



Browse: Recently added



File Tools Metadata Viewers

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

Answer please

This data is about households and their members.

HHID	memberID	c01	c02	c03	c04	c05	c06	c07	c08	c09	c10	c11	c12	c13
101001	1	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	1
101002	1	2	1	25	4	2	4	3	1	0	1	1	1	1
101002	2	1	2	27	4	2	4	3	1	0	1	1	1	1
101002	3	1	3	5					1	0	1	1	1	1
101002	4	2	3	2					1	0	1	1	1	1
101003	1	2	6	75	1	2	5	3	1	2	2	1	1	1
101003	2	2	4	27	1	2	5	3	1	2	2	1	1	1
101003	3	1	3	6					1	2	2	1	1	1
101003	4	2	3	4					1	2	2	1	1	1

Data Re-Use

2D Acoustic Numerical Breast Phantoms and USCT Measurement Data

Version 1.1



Li, Fu; Villa, Umberto; Park, Seonyoung; Anastasio, Mark, 2021, "2D Acoustic Numerical Breast Phantoms and USCT Measurement Data", <https://doi.org/10.7910/DVN/CUFVKE>, Harvard Dataverse, V1

Cite Dataset ▾ Learn about Data Citation Standards.

Access Dataset ▾

Contact Owner Share

Dataset Metrics ⓘ

904 Downloads ⓘ

Description ⓘ

Companion dataset of the manuscript:

Fu Li, Umberto Villa, Seonyoung Park, Mark A. Anastasio, Three-dimensional stochastic numerical breast phantoms for enabling virtual imaging trials of ultrasound computed tomography. *Arxiv preprint 2106.02744* (2021)

This dataset includes a collection of 52 two-dimensional slices of numerical breast phantoms (NBPs) and corresponding ultrasound computed tomography (USCT) simulated measurement data. The anatomical structures of these NBPs were obtained by use of tools from the Virtual Imaging Clinical Center for the development of breast cancer research. <https://doi.org/10.7910/DVN/CUFVKE>

cited by

Generative models based on eigendecomposition for dense ray tracing

The Journal of the Acoustical Society of America 152, 679 (2022); <https://doi.org/10.1121/10.0012973>

Jorge A. Ramos Oliveira, Mario Castelan¹, and Arturo Baltazar

View Affiliations View Contributors



TOPICS

- Covariance and correlation
- Calculus of variations

ABSTRACT

In this wo

REFERENCES

- Li, F., Villa, U., Park, S., and Anastasio, M. (2021). "2D acoustic numerical breast phantoms and USCT measurement data," <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/CUFVKE> (Last viewed June 5, 2022). [Google Scholar](#)

Harvard Dataverse > Integrated Crisis Early Warning System (ICEWS) Dataverse >

ICEWS Coded Event Data

Version 26.0



Boschee, Elizabeth; Lautenschlager, Jennifer; O'Brien, Sean; Shellman, Steve; Starz, James; Ward, Michael, 2015, "ICEWS Coded Event Data", <https://doi.org/10.7910/DVN/28075>, Harvard Dataverse, V36, UNF:6:NOSHB7wyt0SQ8sMg7+w38w== [fileUNF]

Cite Dataset ▾ Learn about Data Citation Standards.

Access Dataset ▾

Contact Owner Share

Dataset Metrics ⓘ

95,873 Downloads ⓘ

Description ⓘ

Event data consists of coded interactions between socio-political actors (i.e., cooperative or hostile actions between individuals, groups, sectors and nation states). Events are automatically identified and extracted from news articles by the BBN ACCENT event coder. These events are essentially triples consisting of a source actor, an event type (according to the CAMEO taxonomy of events), and a target actor. Geographical-temporal metadata are also extracted and associated with the relevant events within a news article. We plan to update this data on a periodic basis. Additional event data may be made available For Official Use Only (FOUO), government sponsored research activities. (2014)

Subject ⓘ

Related Publication ⓘ Shilliday, A., and Lautenschlager, J. Data for a Global ICEWS and Ongoing Research. 2nd International Conference on Peace-Cultural Decision Making. *Event 2015*

cited by



Journal of Asian Economics

Volume 84, February 2023, 101578



The Arab Spring, a setback for gender equality? Evidence from the Gallup World Poll

Robert Rudolf^a ✉, Sh

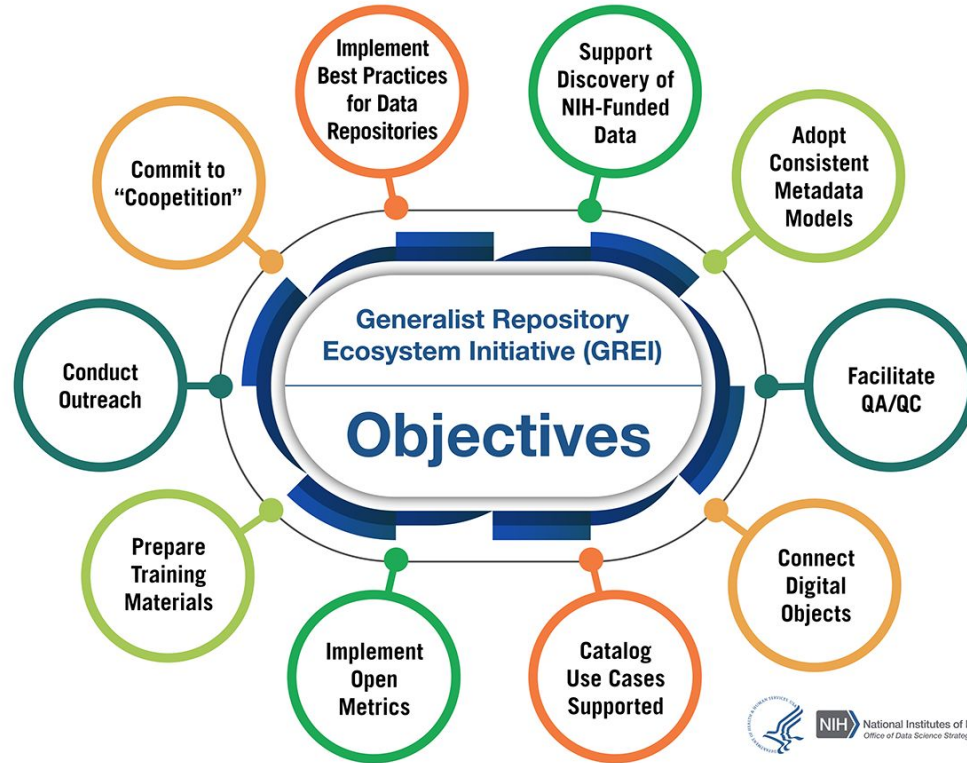
Show more ▾

+ Add to Mendeley

References

- Boschee et al., 2015 Boschee, E., Lautenschlager, J., O'Brien, S., Shellman, J. (2015). *ICEWS coded event data*. Harvard Dataverse, V28, UNF:6:NOSHB7wyt0SQ8sMg7+w38w== [fileUNF]. Retrieved from <https://doi.org/10.7910/DVN/28075> . [Google Scholar](#)

NIH-GREI



Current Features



Installations



Collections



Datasets



Files

NIH OTA Adds...



Datasets

- New Workflows for **Large Data Support** ([Globus](#), [demo](#))
- **Remote** Data Support ([TRSA](#))
- 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 | • Metadata Harvesting |
| • UX/UI Enhancements | • Curation Services |
| • New APIs | • Training and Outreach |
| • Interoperability | |

Summary

- **The Dataverse Project** facilitates the creation of digital repositories to ensure that the data are high quality, properly documented, organized, and accessible.
- **The Harvard Dataverse Repository** is open to the world wide research community for data sharing (free up to 1 terabyte of data, and 2.5GB per file limitation). See our [policies page](#)
- [Metadata](#) is an essential component of the Dataverse Project
- **Best Practices** in a Dataverse repository: Collection creation and customization, dataset creation and file uploads, publishing workflows, and QA to ensure FAIR

NIH Data Management and Sharing Policy: Harvard Dataverse

2023

Data Sharing in Context

“NIH is aware that not just technological advancement, but behavioral change is also necessary to advance data science goals”



Science, Digital; Goodey, Gregory; Hahnel, Mark; Zhou, Yuanchun; Jiang, Lulu; Chandramouliswaran, Ishwar; et al. (2022): The State of Open Data 2022. Digital Science. Report. <https://doi.org/10.6084/m9.figshare.21276984.v5>

NIH Data Management and Sharing Policy

Submission of a Data Management and Sharing Plan (DMSP)

Proposed repository to be used consistent with NIH guidance



Compliance with the awardee's plan as approved by NIH ICO

Elements of a Data Management & Sharing Plan

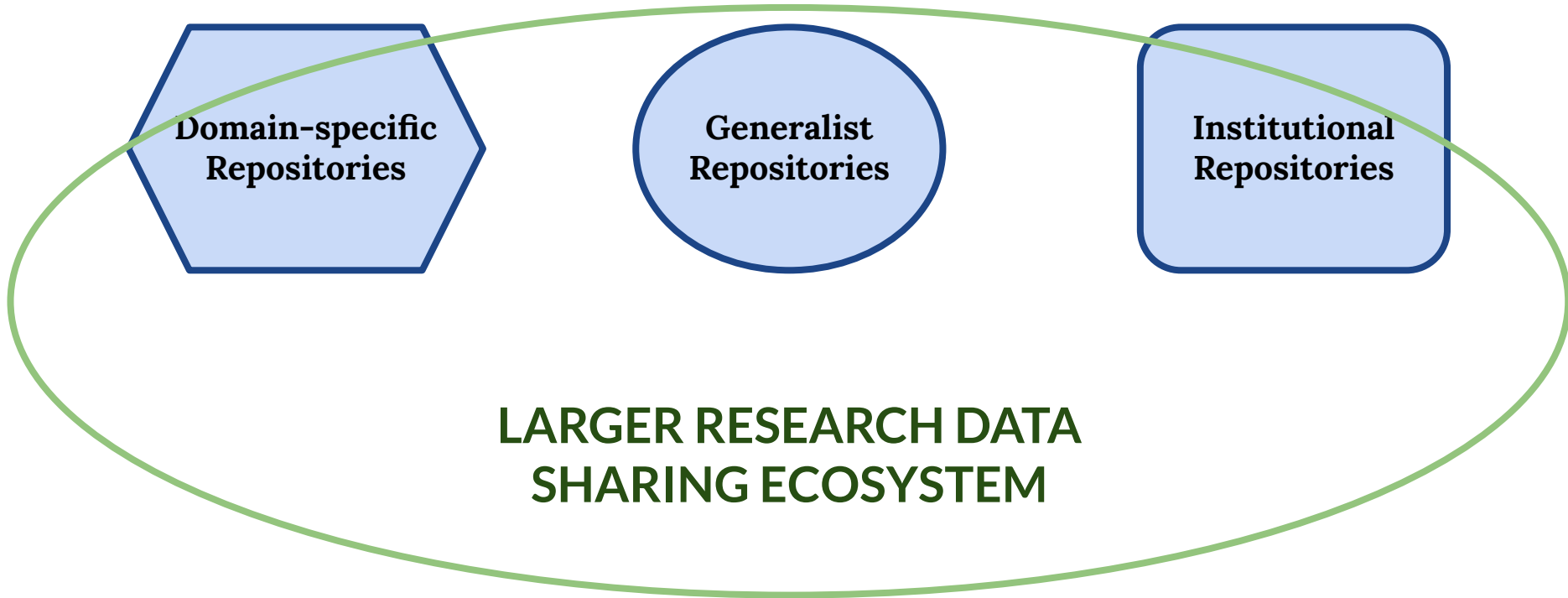
1. **Data Types:** Data to be preserved and shared
2. **Related Tools, Software, Code:** Tools and software needed to access and manipulate data
3. **Common Data Standards:** Standards to be applied to scientific data & metadata
4. **Data Preservation, Access, Timelines:** Repository to be used, persistent unique identifiers, and when/how long data will be available
5. **Access, Distribution, Reuse Considerations:** Factors affecting data access, distribution, or reuse related to informed consent or privacy & confidentiality protections
6. **Oversight of Data Management:** How Plan compliance will be monitored & managed and by whom

NIH Data Repository Desired Characteristics

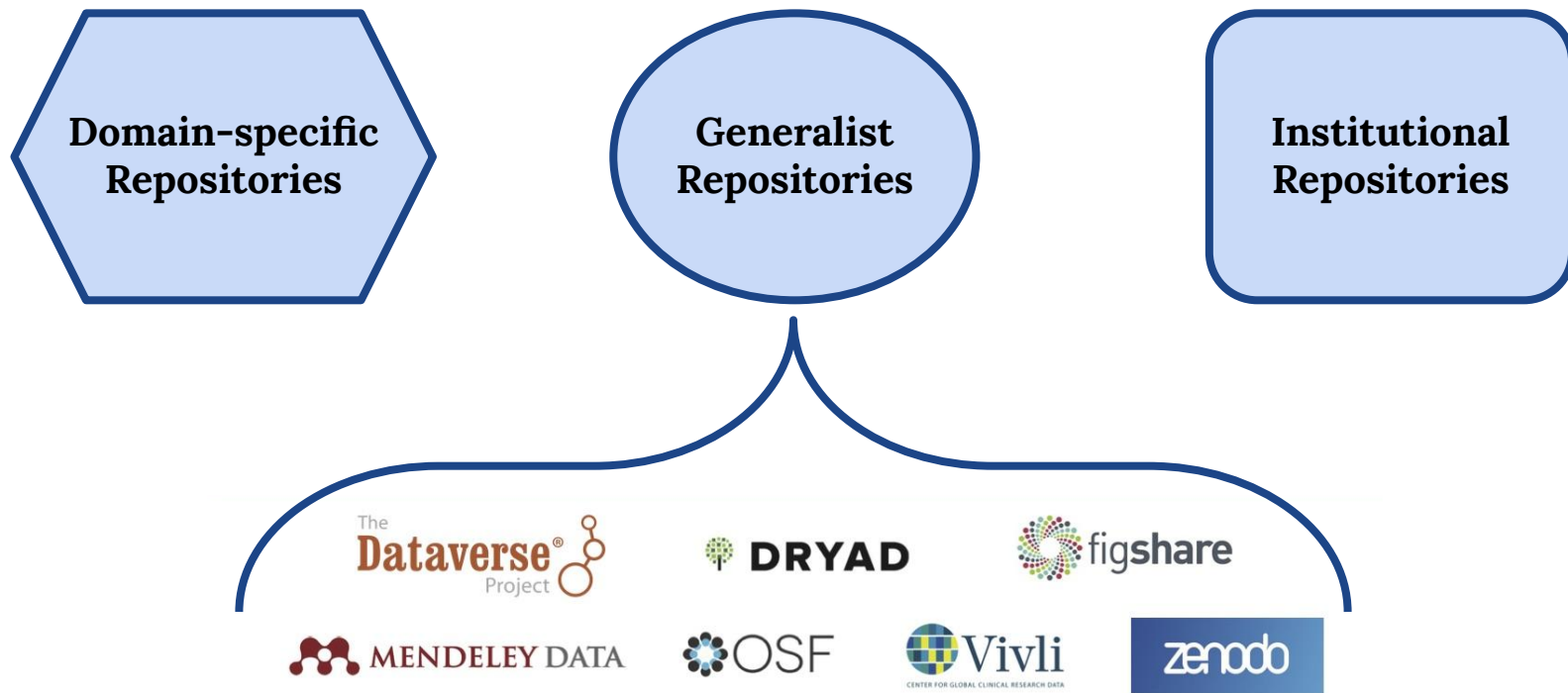
NIH promotes the use of established data repositories because deposit in a quality data repository generally improves the FAIRness (Findable, Accessible, Interoperable, and Re-usable) of the data.

- **Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research:**
<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html>
- **Open Domain-Specific Data Sharing Repositories:**
https://www.nlm.nih.gov/NIHbmic/domain_specific_repositories.html
- **Generalist Repositories:**
https://www.nlm.nih.gov/NIHbmic/generalist_repositories.html

NIH Generalist Repository Ecosystem Initiative



NIH Generalist Repository Ecosystem Initiative






&



HARVARD
Dataverse

Harvard Dataverse NIH-DMP Guidance

<https://support.dataverse.harvard.edu/harvard-dataverse-nih-dmp-guidance>



Getting Started ▾ NIH-GREI ▾ Curation Services Trainings and News ▾ Policies & Governance ▾ About [Go to Harvard Dataverse >](#)

HARVARD DATAVERSE NIH-DMP GUIDANCE

[DMSP Elements](#)

Contact Us


HOME /

Harvard Dataverse NIH-DMP Guidance

Version 1, June 30th, 2023

Introduction

Researchers should consider using a combination of repository features and data sharing best practices to address the elements in the NIH Data Management and Sharing Plan (DMSP). Below are examples and general guidance for describing how Harvard Dataverse Repository addressed some of the questions in the DMSP.

Document licensed under: <https://creativecommons.org/licenses/by-nc-sa/4.0> 

DMSP Elements

- [Element 1: Data Types](#)
- [Element 2: Related Tools, Software and Code](#)
- [Element 3: Standards](#)
- [Element 4: Data Preservation, Access, and Associated Timelines](#)
- [Element 5: Access, Distribution, and Reuse Considerations](#)
- [Element 6: Oversight of Data Management](#)

[Downloadable PDF of the HDV-NIH-DMP Guidance information](#)

Element 1: Data Types

Harvard Dataverse Supports:

- All data types
 - Any data format regardless of file extension
 - Data formats for preservation:
<https://www.openaire.eu/data-formats-preservation-guide>
- Tabular data
 - For SPSS, Stata, RData, and CSV files, the ingest process extracts the data content from the user's files and archives it in an application-neutral, easily-readable format.
 - Note: enhanced ingest for Excel files is not available.

Sample DMP Text: For tabular data: PI will ensure that tabular data files are ingested properly by the Dataverse software. To best support archival preservation, Harvard Dataverse stores the raw data content extracted from successfully ingested tabular data files in plain text, TAB-delimited files. The metadata information that describes this content is stored separately, in a relational database, so that it can be accessed efficiently by the application. For the purposes of archival preservation it can be exported, in plain text XML files, using a standardized, open DDI Codebook format.

Element 2: Related Tools, Software and Code

Harvard Dataverse Recommends:

- Publish analysis code
 - Best practice guidelines encourage the publication of analysis code, configuration, and execution files to be published with data and other input files used to generate or compute research results. This is especially important for replication datasets.
- Use Harvard Dataverse tools
 - Github connector; GeoJSON; Data Explorer; Supported File Previewers; Binder
- Follow best practices for research code
 - Code files - such as Stata, R, MATLAB, or Python files or scripts - have become a frequent addition to the research data deposited in Dataverse repositories.
- Publish research & scientific software
 - Generalized or multi-purpose software that can be used across different datasets may be published as a separate research object

Sample Text: To support data access and reuse all code files generated as a product of this research will be published in Harvard Dataverse as part of the replication dataset that contains data and documentation files.

Original software developed to analyze imaging data (e.g.) will be published in Harvard Dataverse.

Element 3: Standards

Harvard Dataverse Recommends:

- Apply relevant metadata
 - Choose metadata fields from [Harvard Dataverse Supported Metadata Blocks](#)
- Apply additional NIH project metadata
 - Life Sciences Metadata
 - CodeMeta (forthcoming)
- Establish bidirectional link between dataset and related publications
 - Cite related publication/s using related publication field
 - Cite dataset DOI in your publications and reference lists

Sample Text: Harvard Dataverse is committed to using standard-compliant metadata to ensure that metadata can be mapped easily to standard metadata schemas and be exported into JSON format (XML for tabular file metadata) for preservation and interoperability.

The Life Sciences metadata block will be applied to all datasets resulting from this research.

Metadata, configurations, and parameters related specifically to software applications that cannot be described in Dataverse will be recorded in documentation files and will be published along with data.

Element 4: Data Preservation, Access, and Associated Timelines

Harvard Dataverse Policies:

- Repository hosting, backups, and preservation
 - Harvard Library Technical Services (LTS), in collaboration with Harvard University Information Technology (HUIT) and the Institute for Quantitative Social Science (IQSS), hosts the Harvard's Dataverse repository using Amazon Web Services and S3, and maintains a full backup of all data and directories using Amazon Glacier.
 - FAS Research Computing at Harvard University keeps a backup of all Harvard Dataverse data and directories. This means that there are always full, recent copies of the Harvard Dataverse repository at multiple locations.
- Data files access
 - Access to data files is open by default but can be mediated through a variety of features including: Embargos; Restricted file access; Guestbooks; and User permissions.
- Metadata access
 - HDV policy is that published dataset and file level metadata are always openly accessible.

Sample Text: Scientific data from the research/study/project will be published in Harvard Dataverse, a nonprofit, generalist data repository built with the open source Dataverse Project software and affiliated with Harvard University's Institute for Quantitative Social Science and Harvard Library. The Harvard Dataverse Repository is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, to share, archive, cite, access, and explore research data. Data published in Harvard Dataverse are described according to standard metadata schemas, exposed to the web via OAI-PMH and schema.org, assigned DOIs automatically, and backed up regularly to Amazon Glacier and Harvard's Faculty of Arts and Sciences Research Computing servers. Data published in Harvard Dataverse will be accessible to the research community for the long-term.

Element 5: Access, Distribution, and Reuse Considerations

Harvard Dataverse Policies:

- Harvard Dataverse does not support sensitive and personal identifiable data deposits
 - The depositor is responsible for removing all sensitive and personal identifiers for all files uploaded to Harvard repository.
 - Harvard Policy on sensitive data:
<https://support.dataverse.harvard.edu/harvard-dataverse-general-terms-use>
- Access and embargo features, custom terms of access and use available
 - Support includes: file restricted access control via the “request access” feature; file level embargo; depositor-defined terms of use and access
 - Depositor defined access control (example): [Early Head Start Research and Evaluation Project, 1996 - 2001: Childcare/Teacher Data](#)

Sample Text: Data will be published openly under a CC-0 public domain waiver.

Restricted files will be accessible via the “request access” workflow

Additional restricted content in Harvard Dataverse may be accessible via custom terms of access and reuse.

Additional, sensitive data, not published in Harvard Dataverse may be accessible via a formal Data Use Agreement negotiated between research institutions.

Data will be embargoed for 50 years to protect survey participants from potential reidentification.

Element 6: Oversight of Data Management

Harvard Dataverse Policies:

- Role of Harvard Dataverse
 - Harvard Dataverse is not a responsible party that may be named in a Data Management and Sharing Plan. Users retain ownership and control of all deposited data.
- Apply relevant metadata
 - Include depositor information in dataset metadata in “Depositor” field.
- Curation Services can be included in proposed budget
 - The Harvard Dataverse Repository provides fee-based curation services to researchers around the world who are depositing data into The Harvard Dataverse Repository. Research data replication datasets, data for related publications, and all file types and domains are welcomed in the Harvard Dataverse Repository. Our curation services will ensure that your datasets are discoverable, accessible, interoperable, and reusable (FAIR).

Sample Text: FirstName LastName, Job Title, will be responsible for depositing all scientific data resulting from this project in a Harvard Dataverse collection.

We will utilize custom curation services, outlined in our budget justification, to expertly curate datasets to ensure that they are discoverable, accessible, interoperable, and reusable (FAIR).

Budgeting for Harvard Dataverse

- **Curating data**, such as engaging a generalist repository for curation support services
- **Developing supporting documentation**, such as research personnel effort needed to create and retain appropriate documentation beyond the effort needed to simply generate or collect the data
- **Formatting data according to accepted community standards**, or for transmission to and storage at a selected repository for long-term preservation and access, such as support services that may be offered by the repository or engaging a service center for support with data deposits
- **De-identifying data**, such as research personnel effort to systematically remove all identifiers from a data set prior to sharing or engaging a third-party vendor to certify that a data set has been fully de-identified
- **Preparing metadata** to foster discoverability, interpretation, and reuse, such as research personnel effort needed to document appropriate metadata beyond the effort needed to simply generate or collect the data
- **Unique, project-specific information resources** necessary to provide local management and preservation, such as charges from an external vendor for cloud storage prior to deposit into an established repository or costs to use a third-party electronic lab notebook
- **Preserving and sharing data through established repositories**, such as data deposit fees
- **Large data** support requires additional charge for storage/downloads.

Harvard Dataverse Curation Services

<https://support.dataverse.harvard.edu/curation-services>

Service Name	Description	Service Components	Costs	
Data Curation Services			Harvard	Non-Harvard
Free consultation and assessment	Triage level + Office hours Follow-up with recipients of support letters and consultation clients	Demonstrations of software Consult on projects size and scope and fit for Harvard Dataverse	Up to 3 hours	Up to 2 hours
Extended consultation services		<ul style="list-style-type: none"> • Demonstrations of software • Deidentification consult • Dataverse organization consult • File organization consult • Replication verification consult 	\$100/hour	\$200/hour
Dataverse set-up services and dataset and data file ingest	Establish data curation infrastructure and organization enabling data owners to self-curate their data.	<ul style="list-style-type: none"> • Extended Consultation • Dataverse creation and customization • Dataset creation • Data file ingest • Metadata enrichment of dataset and data files • Documentation for use of final product 	\$2000 per "collection" plus \$100 minimum per dataset	\$4000 per "collection" plus \$200 minimum per dataset
Ongoing dataverse administration and curation services	<p>Ongoing maintenance of Dataverse and datasets</p> <p>Continuous updates of dataverse and datasets as needed to reflect software upgrades and sharing standards</p>	<ul style="list-style-type: none"> • Extended consultation services • Dataverse set-up services and dataset and data file ingest • Maintenance of dataverses and datasets within Harvard Dataverse 	\$10,000/year plus \$100 minimum per dataset	\$20,000/year plus \$200 minimum per datasets
Custom services for existing dataverse	<p>One time curation services</p> <p>This includes:</p> <ul style="list-style-type: none"> • file-upload • metadata enhancement (bi-directional linking), • linking affiliated datasets • Documentation • Digitization of audio/paper/content for sharing purposes 		\$100/hour	\$200/hour

Thank You!

The Dataverse Project

<https://dataverse.org/>

The Harvard Dataverse repository

<https://dataverse.harvard.edu/>

Integrating Dataverse and DSpace

<https://osf.io/72w4m>

The Dataverse Guide

<https://guides.dataverse.org/en/latest/>

Dataverse on Github

<https://github.com/IOSS/dataverse>

Dataverse google community

<https://groups.google.com/g/dataverse-community>