

The Dataverse Project

Training for Barcelona SuperComputing Center,
September 13th, 2024

The Team



Sonia Barbosa
*Associate Director, Dataverse, Data
Curation, and The Murray Archive*



Ceilyn Boyd
*Dataverse Development Project
Manager*



Gustavo Durand
*Dataverse Technical Lead &
Architect*



Philip Durbin
Software Developer



Julian Gautier
Product Research Specialist

Agenda:

- Dataverse Overview
- Software technology & features
- Metadata
- Repository Management and Data Curation
- Dataverse Community

Dataverse Overview

Dataverse Project

About the Dataverse Project

- An open-source repository to publish, cite, and archive research data
- Built to support multiple types of data, users, and workflows
- Developed at Harvard's Institute for Quantitative Social Science (IQSS)
- Established in 1997, Dataverse Network Launch in 2006
- Development funded by IQSS and with grants, in collaboration with institutions around the world
- 2022 - present: Re-Architecture
- Core team
 - @ IQSS - developers, UX/UI, metadata specialists, curation team, leadership team
 - Key contributors from the community with full privileges as IQSS team

Dataverse Project Goals

- Grow the Dataverse community
- Empower the open source community to explore and implement new Dataverse applications, tools, and services
- Develop the capability to handle sensitive data and big data
- Expand data and metadata features for existing and new disciplines
- Expand archival and preservation features
- Increase interoperability through the implementation of standards
- Increase contributions from the open source development community
- Improve the Dataverse user experience

King, G. (2007). An Introduction to the Dataverse Network as an infrastructure for data sharing. *Sociological Methods & Research*, 36(2), 173–199.

Altman & King (2007). A Proposed Standard for the Scholarly Citation of Quantitative Data, *D-Lib Magazine*, 13, 3/4(March/April, 2007).

Wilkinson, et al. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3.

1997

Project Begins
Focus on preserving & sharing social sciences data
Harvard-MIT Data Center collaboration
Murray Data Archive

2006

The Dataverse Network 1.0
Harvard Dataverse Repository established


2007

 Announcing the Dataverse Network

2008

Odum Archive becomes first Dataverse outside Harvard

2015

Data Citations
 First Open Dataverse Community Meeting
Dataverse 4.0

2016

FAIR data

2023

 100th Dataverse installation
Dataverse 6.0
Dataverse Re-architecture
 AI comes to Harvard Dataverse

2024

 The Next Frontier!
Large data
Computing on data

Altman, Borgman, Crosas & Matone. (2015). An introduction to the joint principles for data citation. *Bulletin of the Association for Information Science and Technology*, 41(3), 43–45.

Technology Evolution

Prehistory: VDC

- Written using XSLT (XSL Transformations)

History: Dataverse Network

- v1.0 - first published version using Java (Glassfish Web Server) for both backend and frontend
- v3.0 - initial implementation of APIs for accessing DVN data

Today: Dataverse

- v4.0 - Complete rewrite of backend and frontend
 - Flexibility - Dataverse Collection Structure, Metadata, Permissions
 - APIs and External Tools (added in v4.8.5)
- v5.0 - Infrastructure upgrade from Glassfish 4 to Payara 5
- v6.0 - Latest infrastructure upgrade to Payara 6

Future

- SPA - Frontend written in React and communicating exclusively via APIs with backend
- Marketplace - for easier configuration of modular components - exporters, metadata blocks, etc.
- Dataverse network metrics hub
- More Integrations and External tools!

Dataverse Technology & Features

Dataverse Features

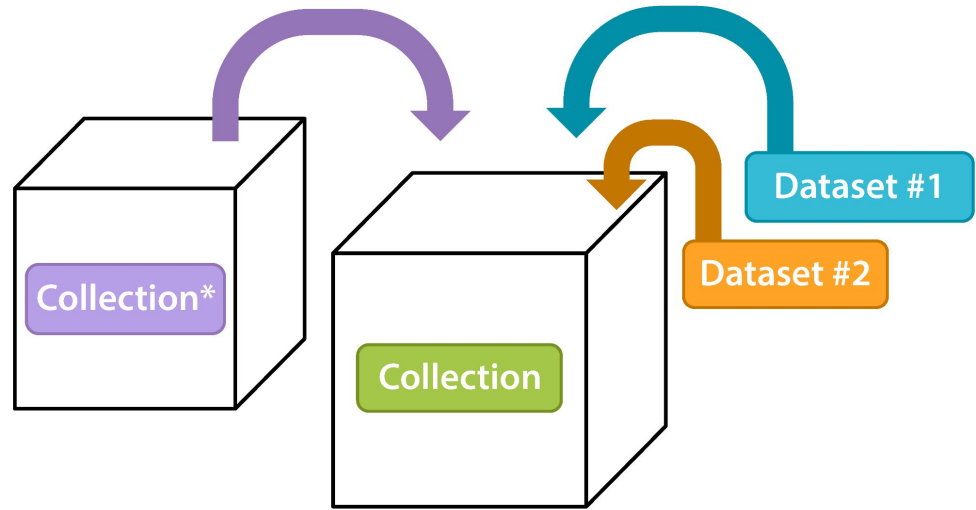
<https://dataverse.org/software-features>

- Main goal of core code is to focus on publishing (citing, sharing, versioning, etc.), FAIR Data principles
- Robust APIs to allow interoperability with “external tools” and other repositories / software

Dataverse Collections

- Ability to create Dataverse collections to organize datasets according to your needs
- Dataverses collections can also contain other collections, enabling any hierarchical structure
- Different rules can be applied for different Dataverse collections, e.g. for Metadata, Permissions, etc.

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



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

* Collections can contain other Collections

Dynamic Metadata

- Metadata is defined dynamically at the database level, allowing for modularly adding new Metadata blocks
- Supports:
 - single or multiple values
 - simple or compound values
 - controlled vocabularies

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](#)

Citation Metadata ▲

Title * ⓘ
[Add "Replication Data for" to Title](#)

Author * ⓘ

Name * ⓘ **Affiliation** ⓘ +

Identifier Scheme ⓘ **Identifier** ⓘ

Contact * ⓘ

Name ⓘ **Affiliation** ⓘ +

E-mail * ⓘ

Description * ⓘ This field supports only certain HTML tags.

Text * ⓘ

Date ⓘ

Flexible Permission System

- Supports multiple workflows by controlling who can add to your Dataverse collection, what they can, and what role they have on and created Datasets
- Roles are defined as a set of permissions to grant to users or to groups
- Groups can be defined statically or dynamically (e.g. users logging in from the same institution, via Shibboleth)

Edit Access

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

When a user adds a new dataset to this dataverse, which role should be automatically assigned to them on that dataset?

- 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

2 Users/Groups

User/Group Name (Affiliation) ⚙	ID ⚙	Role ⚙
Dataverse Admin (Dataverse.org)	@dataverseAdmin	Admin
Anyone with a Dataverse account	:authenticated-users	Dataverse + Dataset Creator

Robust APIs

- APIs for search, deposit, access, administration, metrics, etc.
- Additional APIs for harvesting (discovery) and interoperability with other systems
- External tools can be registered via APIs, so that Dataverse can provide links in the UI, then user is sent to tool to preview, explore, configure, and more

API Guide

Contents:

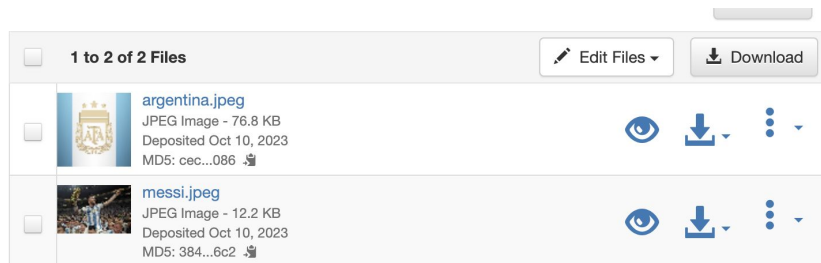
- Introduction
 - What is an API?
 - Types of Dataverse Software API Users
 - API Users Within a Single Dataverse Installation
 - Users of Integrations and Apps
 - Power Users
 - Support Teams and Superusers
 - Sysadmins
 - In House Developers
 - API Users Across the Dataverse Project
 - Developers of Integrations, External Tools, and Apps
 - Developers of Dataverse Software API Client Libraries
 - Developers of The Dataverse Software Itself
 - How This Guide is Organized
 - Getting Started
 - API Tokens and Authentication
 - Lists of Dataverse APIs
 - Client Libraries
 - Examples
 - Frequently Asked Questions
 - Getting Help
- Getting Started with APIs
 - Servers You Can Test With
 - Getting an API Token
 - curl Examples and Environment Variables
 - Depositing Data
 - Creating a Dataverse Collection
 - Creating a Dataset
 - Uploading Files
 - Publishing a Dataverse Collection
 - Publishing a Dataset

Dataverse Community Integrations & Tools

- Tools that talk to Dataverse
 - generally used to deposit data into Dataverse (via Deposit API)
 - usually don't require anything special to be set up in the Dataverse repository
- Tools that Dataverse talks to
 - user starts on Dataverse and is directed to the external tool
 - require manifest files
 - have predefined areas in the UI where these would plug into (**Explore** tools)
 - **OR**, are embedded into the Dataverse UI directly (**Preview** tools and **Query** Tools)
- Tools that do both
 - user starts on Dataverse and is directed to the external tool
 - require manifest files
 - also have predefined areas in the UI where these would plug into (**Configure** tools)
 - will also send something back to Dataverse, so need an API token that has "write" privileges









File Previewers

- A set of tools that display the content of files, allowing them to be viewed without downloading the file, including
- audio
- html
- Hypothes.is annotations
- images
- PDF
- text
- video
- tabular data
- spreadsheets
- GeoJSON
- Zip files
- NcML files
- Previewers are available through the preview (eye) icon on Dataset pages
- And also embedded as a tab on Datafile pages



1 to 2 of 2 Files

Edit Files ▾ Download

<input type="checkbox"/>		argentina.jpeg JPEG Image - 76.8 KB Deposited Oct 10, 2023 MD5: cec...086	  
<input type="checkbox"/>		messi.jpeg JPEG Image - 12.2 KB Deposited Oct 10, 2023 MD5: 384...6c2	  

Preview

Metadata

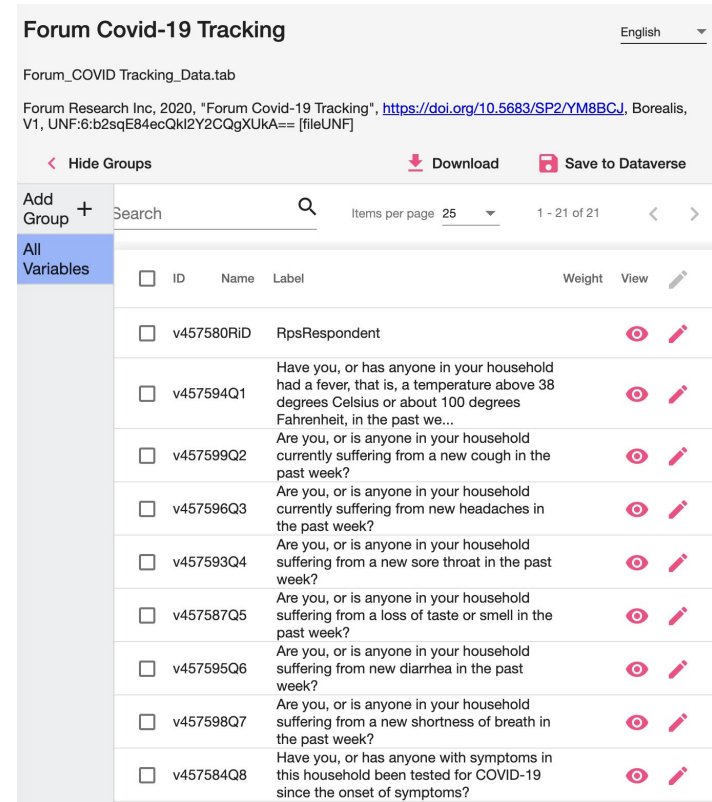
Versions

Explore on View Image





















File Exploration, Configuration, and Query Tools

- File level **explore** tools provide a variety of features from data visualization to statistical analysis
- File level **query** tools allow the user to ask questions (e.g. natural language queries) of a data table's contents without having to download the file
- File level **configure** tools allow (authorized) users to send metadata about the file back to Dataverse

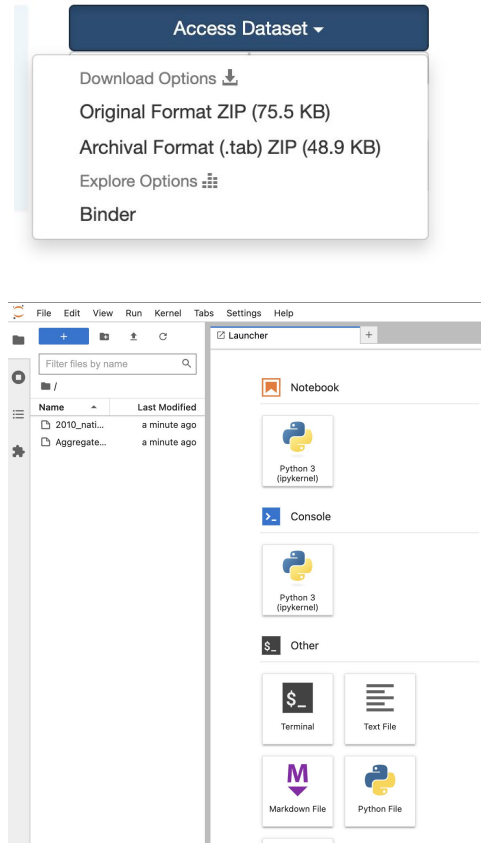


The screenshot displays the 'Forum Covid-19 Tracking' interface. At the top, it shows the title 'Forum Covid-19 Tracking' and the language 'English'. Below this, the file name 'Forum_COVID Tracking_Data.tab' is listed, along with a copyright notice for Forum Research Inc. (2020) and a DOI link. Action buttons for 'Hide Groups', 'Download', and 'Save to Dataverse' are visible. The main area shows a table of variables with columns for 'ID', 'Name', 'Label', 'Weight', and 'View'. The 'All Variables' tab is selected, and the table lists several variables related to COVID-19 symptoms and household characteristics.

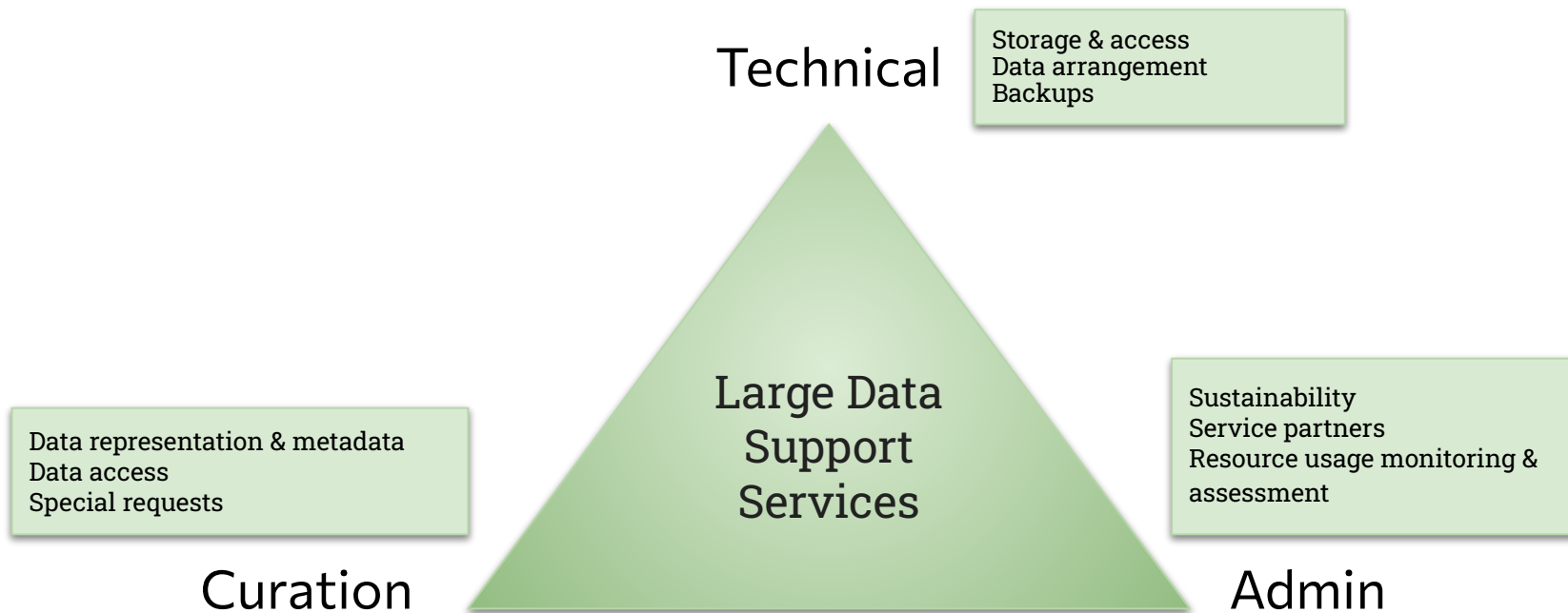
<input type="checkbox"/>	ID	Name	Label	Weight	View	
<input type="checkbox"/>	v457580RID	RpsRespondent				 
<input type="checkbox"/>	v457594Q1		Have you, or has anyone in your household had a fever, that is, a temperature above 38 degrees Celsius or about 100 degrees Fahrenheit, in the past we...			 
<input type="checkbox"/>	v457599Q2		Are you, or is anyone in your household currently suffering from a new cough in the past week?			 
<input type="checkbox"/>	v457596Q3		Are you, or is anyone in your household currently suffering from new headaches in the past week?			 
<input type="checkbox"/>	v457593Q4		Are you, or is anyone in your household suffering from a new sore throat in the past week?			 
<input type="checkbox"/>	v457587Q5		Are you, or is anyone in your household suffering from a loss of taste or smell in the past week?			 
<input type="checkbox"/>	v457595Q6		Are you, or is anyone in your household suffering from new diarrhea in the past week?			 
<input type="checkbox"/>	v457598Q7		Are you, or is anyone in your household suffering from a new shortness of breath in the past week?			 
<input type="checkbox"/>	v457584Q8		Have you, or has anyone with symptoms in this household been tested for COVID-19 since the onset of symptoms?			 

Dataset External Tools

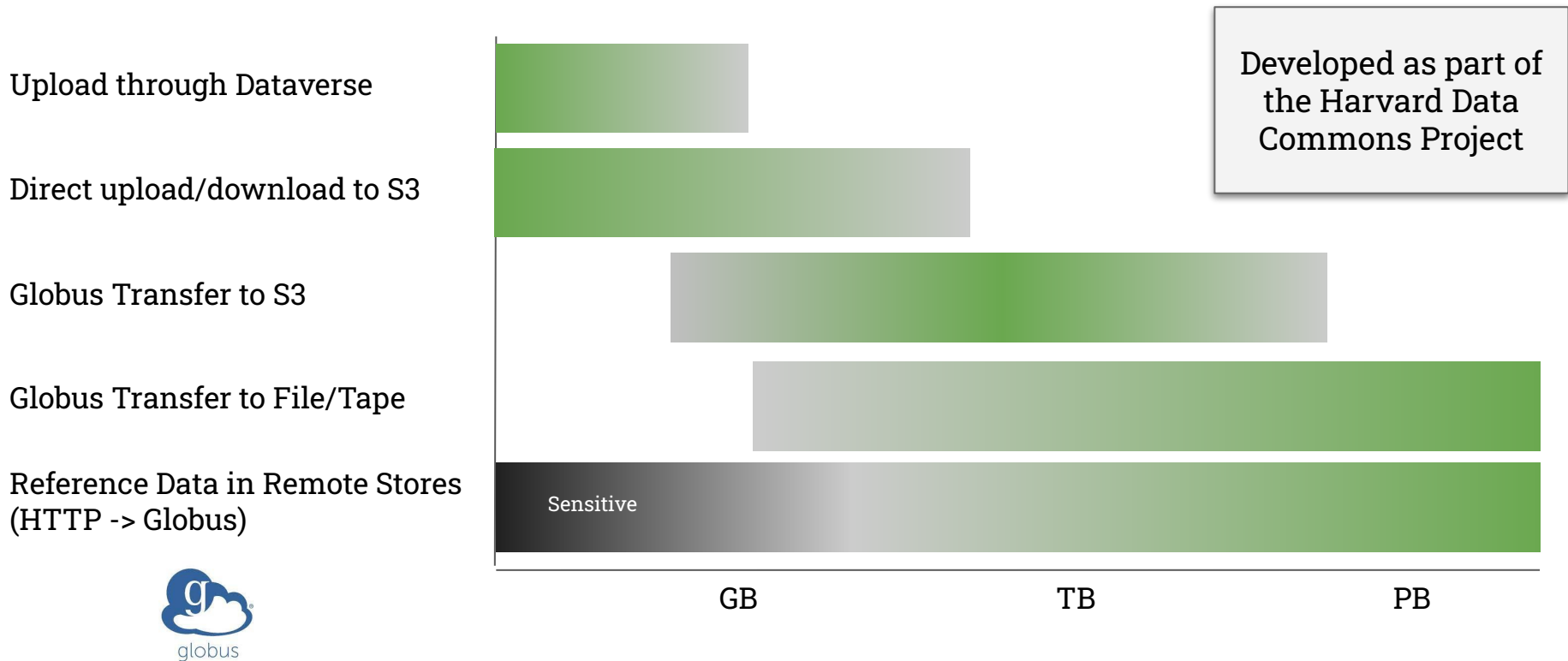
- Dataset level **explore** tools allow the user to explore all the files in a dataset - common use case is reproducibility
 - **WholeTale** - creates reproducible research packages based on popular tools such as Jupyter and RStudio
 - **Binder** - spins up custom computing environments in the cloud (including Jupyter notebooks)
- Dataset level **configure** tools allow (authorized) users to send metadata about the dataset back to Dataverse
 - **Turbo Curator** - uses Open AI's ChatGPT & ICPSR best practices to provide recommendation to enhance metadata & generate meaningful titles, descriptions, and keywords



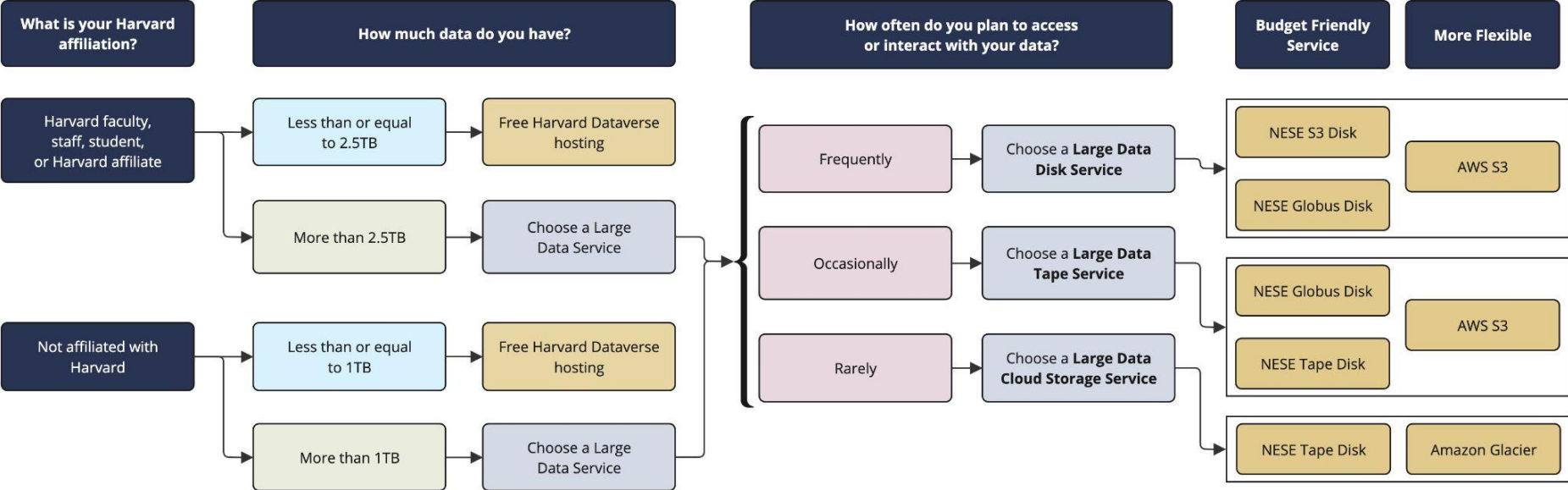
Building and delivering sustainable large data services: challenges & considerations



Harvard Dataverse large data upload & access



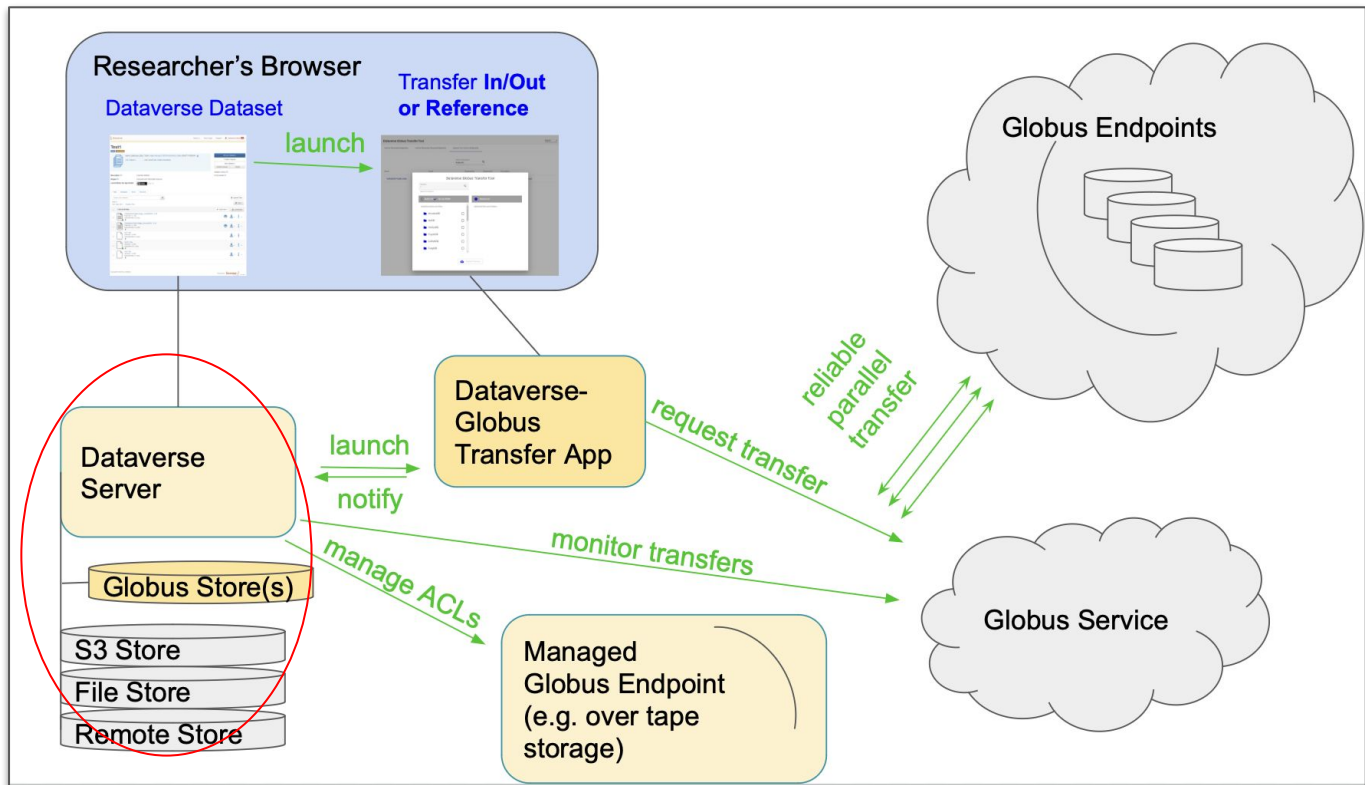
Harvard Dataverse Large Data Services



Computing on Harvard Dataverse datasets (experimental)

Previewers and AI tools to interrogate data are integrated into the Dataverse UI but work mostly for small data.

So far, large data must be downloaded over the net in order to enable computing.



Dataverse Metadata

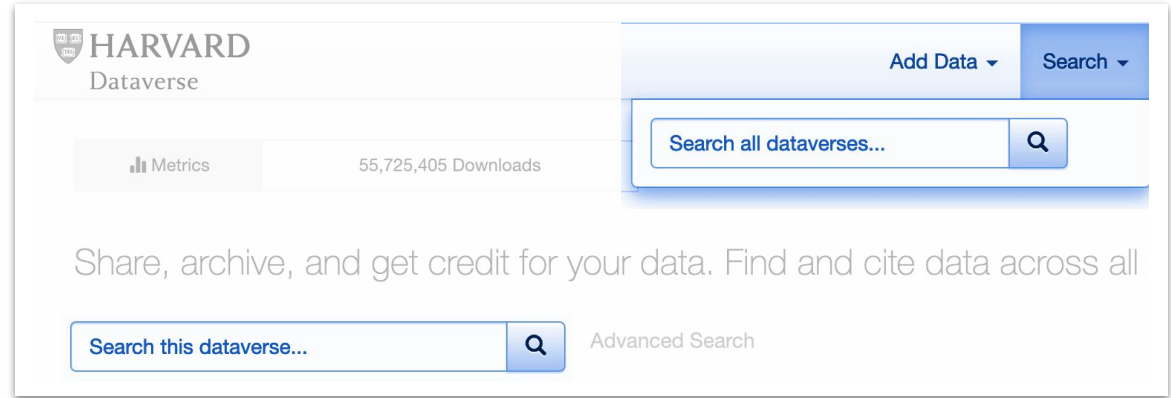
1. An overview of metadata fields in Dataverse
2. How metadata helps others find, use data
3. Enabling and editing metadata fields

1. An overview of metadata fields in Dataverse

	Entered by depositors	Generated by Dataverse software
Collections	Name, URL, description, category, affiliation, contact emails	Creation and publication dates
Datasets	In “metadata blocks” Titles; names and affiliations of authors and contributors; different types of dates and much more Licenses and other types of terms of use and data use agreements; and much more	Persistent IDs, dates, version numbers, use metrics, checksums, differences between versions
Files	File names, description, file paths, tags, restrictions, embargoes	File names, file paths, persistent IDs, file types, sizes, names, data extracted from tabular data and FITS files, checksums, download counts, dates

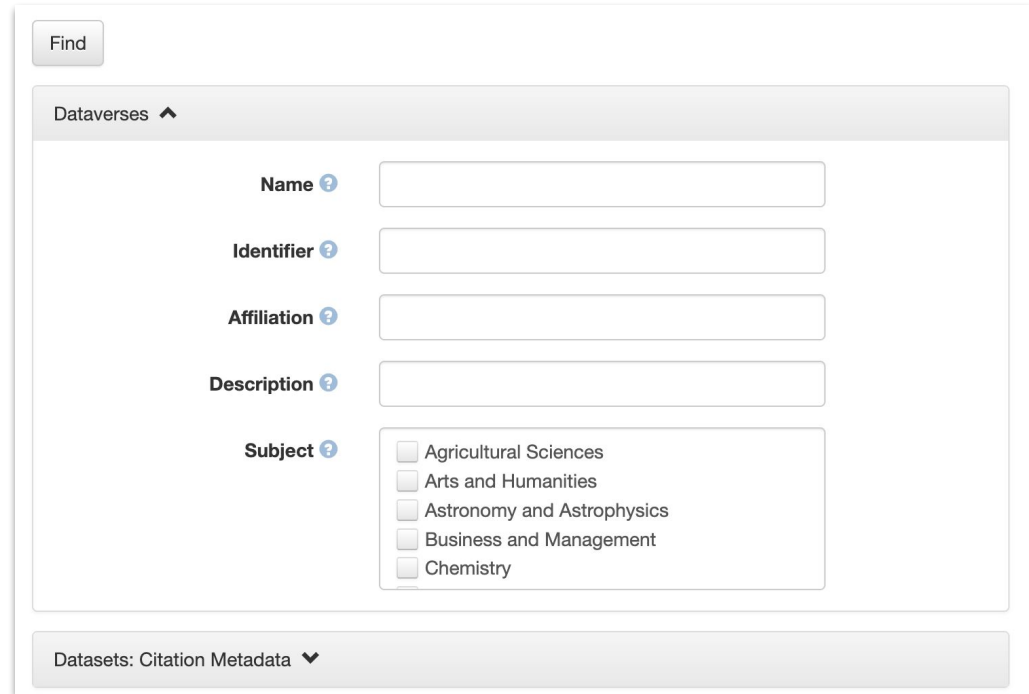
2. How metadata helps others find, use data

- Searching in a Dataverse repository
- Searching outside of a Dataverse repository
- Tools for exploring and editing data



2. How metadata helps others find, use data

- Searching in a Dataverse repository
- Searching outside of a Dataverse repository
- Tools for exploring and editing data



The image shows a search interface for a Dataverse repository. At the top left is a "Find" button. Below it is a header "Dataverses" with an upward-pointing arrow. The main search area contains five fields, each with a label and a question mark icon:

- Name**: A text input field.
- Identifier**: A text input field.
- Affiliation**: A text input field.
- Description**: A text input field.
- Subject**: A list of subjects with checkboxes:
 - Agricultural Sciences
 - Arts and Humanities
 - Astronomy and Astrophysics
 - Business and Management
 - Chemistry

At the bottom of the interface is a footer bar with the text "Datasets: Citation Metadata" and a downward-pointing arrow.

2. How metadata helps others find, use data

- Searching in a Dataverse repository
- Searching outside of a Dataverse repository
- Tools for exploring and editing data

The screenshot shows a search results page for 'Dataverses (6,345)', 'Datasets (152,193)', and 'Files (2,061,989)'. The results are sorted by 'Name (A-Z)'. The first result is 'Replication Data for "You Had Better Mention All of Them:" Race and Gender Effects in Election Loss Narratives' by Haines, Pavielle, 2023. The second result is 'First-choice assay - 12 replicates' by Costa-da-Silva, AL, 2023. The third result is '1-Time-lapsed First-choice assay (Florida International University)' by Andre Silva Dataverse. The fourth result is 'Oviseekmeter assay (choice dynamic) - 30 gravid females - 9 replicates' by Costa-da-Silva, AL, 2023. The page also includes filters for 'Dataverse Category', 'Metadata Source', 'Publication Year', and 'Subject'.

Dataverses (6,345)
Datasets (152,193)
Files (2,061,989)

Dataverse Category
Research Project (2,356)
Researcher (1,981)
Organization or Institution (504)
Research Group (444)
Journal (130)
[More...](#)

Metadata Source
Harvard Dataverse (95,091)
Harvested (63,447)

Publication Year
2023 (7,513)
2022 (19,832)
2021 (23,904)
2020 (9,905)
2019 (4,510)
[More...](#)

Subject
Social Sciences (61,213)
Arts and Humanities (36,231)
Medicine, Health and Life Sciences (9,478)

1 to 10 of 158,538 Results Sort

Replication Data for "You Had Better Mention All of Them:" Race and Gender Effects in Election Loss Narratives"
Nov 13, 2023
Haines, Pavielle, 2023, "Replication Data for "You Had Better Mention All of Them:" Effects in Election Loss Narratives", <https://doi.org/10.7910/DVN/FDFMZT>, Harvard Dataverse, V1
UNF:6:8RX4GeZTwMdhjmZSMadPOw== [fileUNF]
The replication files posted here correspond to the article Replication Data for "You Had Better Mention All of Them:" Race and Gender Effects in Election Loss Narratives." This article is available through Political Research Quarterly.

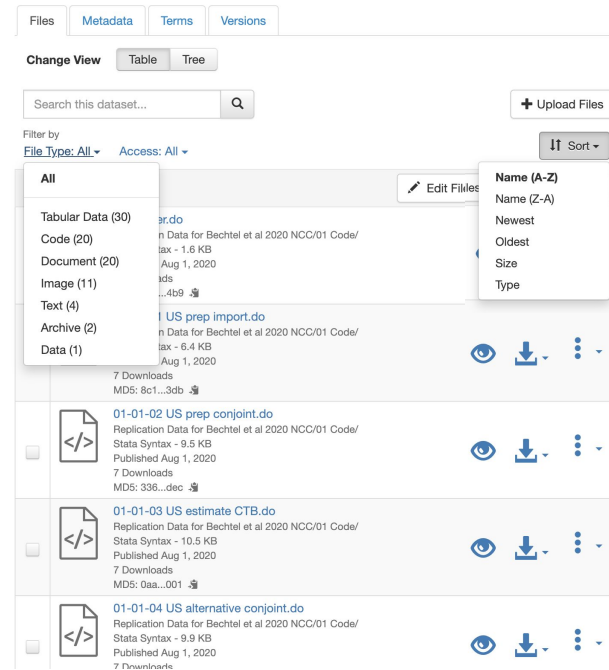
First-choice assay - 12 replicates
Nov 13, 2023 - 1-Time-lapsed First-choice assay
Costa-da-Silva, AL, 2023, "First-choice assay - 12 replicates", <https://doi.org/10.7910/DVN/69COMS>, Harvard Dataverse, V1
These are the time-lapse recorders for 12 trials in the 2-choice water trap assay. Each trial has multiple files in order. The files are named by the number of the experiment (ExpX), replicates on the same day (lowercase letters, when there is more than one trial on the same day,...

1-Time-lapsed First-choice assay (Florida International University)
Nov 13, 2023 Andre Silva Dataverse

Oviseekmeter assay (choice dynamic) - 30 gravid females - 9 replicates
Nov 13, 2023 - 2-Time-lapsed Oviseekmeter assay
Costa-da-Silva, AL, 2023, "Oviseekmeter assay (choice dynamic) - 30 gravid females - 9 replicates", <https://doi.org/10.7910/DVN/BZY3YJ>, Harvard Dataverse, V1

2. How metadata helps others find, use data

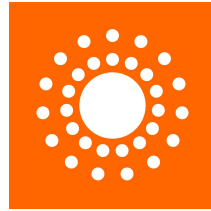
- Searching in a Dataverse repository
- Searching outside of a Dataverse repository
- Tools for exploring and editing data



The screenshot displays the Dataverse user interface. At the top, there are tabs for 'Files', 'Metadata', 'Terms', and 'Versions'. Below these, there are 'Change View' buttons for 'Table' and 'Tree'. A search bar is present with the placeholder text 'Search this dataset...'. To the right of the search bar is an 'Upload Files' button. Below the search bar, there are filters for 'File Type: All' and 'Access: All'. A 'Sort' dropdown menu is open, showing options: 'Name (A-Z)', 'Name (Z-A)', 'Newest', 'Oldest', 'Size', and 'Type'. The main content area shows a list of data files. The first file is '01-01-02 US prep conjoint.do', which is a Stata Syntax file (9.5 KB) published on Aug 1, 2020. It has 7 downloads and an MD5 hash of 8c1...3db. The second file is '01-01-03 US estimate CTB.do', also a Stata Syntax file (10.5 KB) published on Aug 1, 2020, with 7 downloads and an MD5 hash of 0aa...001. The third file is '01-01-04 US alternative conjoint.do', a Stata Syntax file (9.9 KB) published on Aug 1, 2020, with 7 downloads. Each file entry includes a file icon, a title, a description, file size, publication date, download count, and MD5 hash.

2. How metadata helps others find, use data

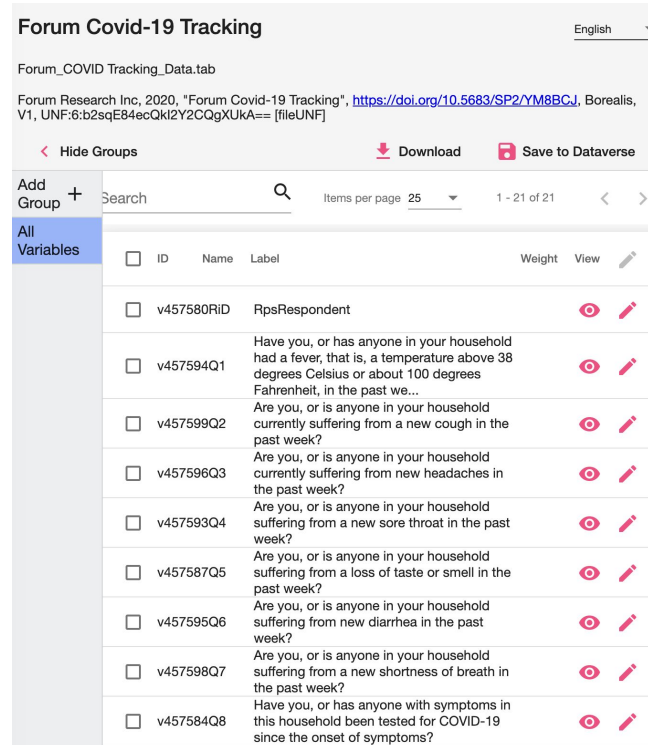
- Searching in a Dataverse repository
- Searching outside of a Dataverse repository
- Tools for exploring and editing data



[Dataverse Metadata Crosswalk](#)

2. How metadata helps others find, use data

- Searching in a Dataverse repository
- Searching outside of a Dataverse repository
- Tools for exploring and editing data



The screenshot displays the Dataverse interface for the dataset 'Forum Covid-19 Tracking'. The page title is 'Forum Covid-19 Tracking' with a language dropdown set to 'English'. Below the title, the dataset name 'Forum_COVID Tracking_Data.tab' is shown, followed by the citation: 'Forum Research Inc, 2020, "Forum Covid-19 Tracking", <https://doi.org/10.5683/SP2/YM8BCJ>, Borealis, V1, UNF:6:b2sqE84ecQkI2Y2CCqgXUKA== [fileUNF]'. Action buttons for 'Hide Groups', 'Download', and 'Save to Dataverse' are visible. A search bar is present with 'Items per page' set to 25 and '1 - 21 of 21' items displayed. A sidebar on the left shows 'All Variables' selected. The main content is a table of variables:

<input type="checkbox"/>	ID	Name	Label	Weight	View	
<input type="checkbox"/>	v457580RiD	RpsRespondent				
<input type="checkbox"/>	v457594Q1		Have you, or has anyone in your household had a fever, that is, a temperature above 38 degrees Celsius or about 100 degrees Fahrenheit, in the past we...			
<input type="checkbox"/>	v457599Q2		Are you, or is anyone in your household currently suffering from a new cough in the past week?			
<input type="checkbox"/>	v457596Q3		Are you, or is anyone in your household currently suffering from new headaches in the past week?			
<input type="checkbox"/>	v457593Q4		Are you, or is anyone in your household suffering from a new sore throat in the past week?			
<input type="checkbox"/>	v457587Q5		Are you, or is anyone in your household suffering from a loss of taste or smell in the past week?			
<input type="checkbox"/>	v457595Q6		Are you, or is anyone in your household suffering from new diarrhea in the past week?			
<input type="checkbox"/>	v457598Q7		Are you, or is anyone in your household suffering from a new shortness of breath in the past week?			
<input type="checkbox"/>	v457584Q8		Have you, or has anyone with symptoms in this household been tested for COVID-19 since the onset of symptoms?			

3. Enabling and editing metadata fields

- Enabling and editing fields
- More editing options
- Guidelines for creating new fields

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](#)
- Computational Workflow Metadata [\[+\] View fields](#)

3. Enabling and editing metadata fields

- Enabling and editing fields
- More editing options
- Guidelines for creating new fields

Metadata

*Asterisks indicate required fields

Citation Metadata ▼

Geospatial Metadata ▼

Social Science and Humanities Metadata ▼

Astronomy and Astrophysics Metadata ▼

Life Sciences Metadata ▼

Journal Metadata ▼

3. Enabling and editing metadata fields

- Enabling and editing fields
- More editing options
- Guidelines for creating new fields

Metadata Customization Guide

	A	B	C	D	E	F	G
1	#metadataBlock	name	dataverseAlias	displayName	blockURI		
2		citation		Citation Metadata	https://dataverse.org		
3	#datasetField	name	title	description	watermark	fieldType	displayOrder
4		title	Title	The main title of the Dataset		text	0
5		subtitle	Subtitle	A secondary title that amplifies or states certain limitations on the main title		text	1
6		alternativeTitle	Alternative Title	Either 1) a title commonly used to refer to the Dataset or 2) an abbreviation of the main title		text	2
7		alternativeURL	Alternative URL	Another URL where one can view or access the data in the Dataset, e.g. a project or personal webpage	https://	url	3
8		otherId	Other Identifier	Another unique identifier for the Dataset (e.g. producer's or another repository's identifier)		none	4
9		otherIdAgency	Agency	The name of the agency that generated the other identifier		text	5
10		otherIdValue	Identifier	Another identifier uniquely identifies the Dataset		text	6
11		author	Author	The entity, e.g. a person or organization, that created the Dataset		none	7
12		authorName	Name	The name of the author, such as the person's name or the name of an organization	1) Family Name, Given Name or 2) Organization XYZ	text	8
13		authorAffiliation	Affiliation	The name of the entity affiliated with the author, e.g. an organization's name	Organization XYZ	text	9
14		authorIdentifier	Identifier Type	Uniquely identifies the author when paired with an identifier type		text	11
15		authorIdentifierScheme	Identifier	The type of identifier that uniquely identifies the author (e.g. ORCID, ISNI)		text	10

3. Enabling and editing metadata fields

- Enabling and editing fields
- More editing options
- Guidelines for creating new fields
 - Consider existing fields
 - Consider discovery in other systems
 - Automate metadata entry
 - Evaluate success
 - Review [metadata text guidelines](#)
 - Keep number of child fields low
 - Avoid deeply nested fields

Data Curation with Dataverse

Topics:

- Repository Management
- Data Curation
- Harvard Dataverse Examples
- Additional user considerations
- Collections & Datasets

Data Repository Management and Curation

Planning/Setup

1. **Repository Goals**
2. **Policies**

Data Acquisition

1. **Data Ingestion**
2. **Metadata Management**

Data Curation

1. **Quality Control**
2. **File organization**
3. **Documentation**

Storage and Preservation

1. **Backup and Redundancy**
2. **Data Formats**
3. **Preservation Plan**

Access/Distribution

1. **Data Access Policies**
2. **Licensing**
3. **Persistent Identifiers**

User Support/Outreach

1. **Training/Documentation**
2. **Community Engagement**

Monitoring/Evaluation

1. **Usage Analytics**
2. **Feedback Mechanism**

Sustainability

1. **Funding models**
2. **Partnerships**

Repository Management: Key Elements

Metadata Standards

Access control and licensing

Storage and backup strategies

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](#)
- Computational Workflow 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

Available Fields	Selected
Author Affiliation	Subject
Language	Author Name
Producer Name	Keyword Term
Production Date	Topic Classification Term
Production Location	
Contributor Type	
Contributor Name	
Funding Information Agency	

Request Access

GSP_list_140630.csv
Plain Text - 636.4 KB
Published Aug 24, 2014
903 Downloads
MD5: 27e...4a8

Demographic, cognitive/behav

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.

Terms of Use
Custom Dataset Terms — the following Custom Dataset Terms of Use take precedence over Network Terms of Use

I request access to data collected as part of the Brain On Massachusetts General Hospital, and I agree to the following:
1. I will not attempt to establish the identity of or attempt to link any of the distributed data to human subjects.
2. I understand that under no circumstances will the code to me, nor will any additional information about individual data for

- 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

- LTS
- HUIT
- Globus
- NESE
- Other...

Data Curation: Key Elements

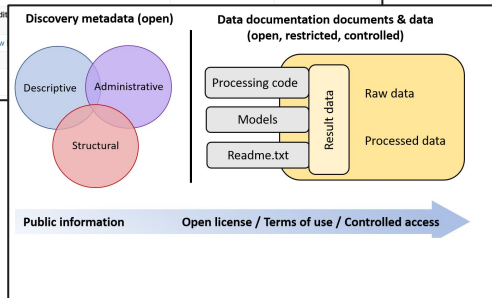
FAIR Principles

Data documentation and version control

Data Sharing Policies



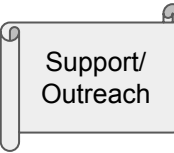
Dataset Version	Summary	Contributors	Published on
4.0	Files (Added: 37); View Details	Dwayne Liburd, Karla Cruz Sanchez	2024-09-12
3.3	Files (Changed File Metadata: 6); View Details	Sonia Barbosa	2023-10-05
3.2	Citation Metadata: Author (1 Changed); Description (1 Changed); Additional Citation Metadata: (1 Changed); MIRA Metadata: (2 Added); Terms of Use/Access Changed View Details	Michelle Rydér, Dwayne Liburd	2018-02-13
3.1	Citation Metadata: Description (1 Changed); Addit of Use/Access Changed View Details		
3.0	Citation Metadata: Description (1 Changed); View Details		
2.0	Terms of Use/Access Changed View Details		
1.0	This is the first published version.		



- Policy Framework
- Access control
- Licensing

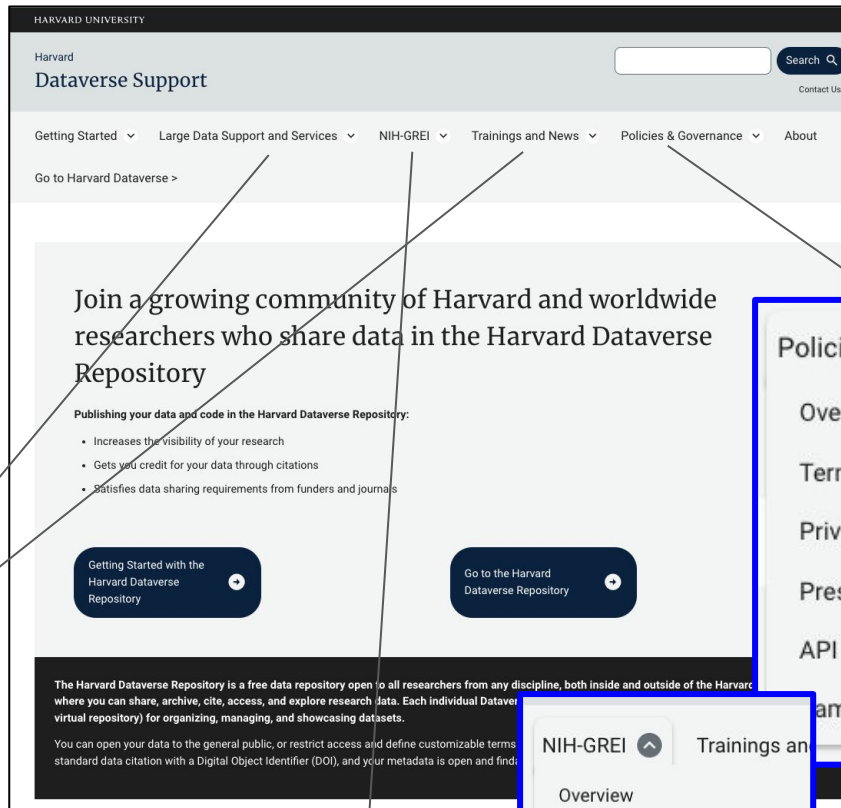
Policies/Terms of use

- Repository Policy Page
- Terms of Use
- Account creation terms of use
- Data use terms of use



Support/
Outreach

- Virtual Introduction to Dataverse APIs**
Jan 21, 2023 | 1:00PM - 2:00PM | Online via Zoom
While Dataverse supports file uploads through its standard web interface, there are a variety of big data use cases where the Dataverse API and tools that use it can simplify the upload process. @tkaplan@fas.harvard.edu
- Virtual Intermediate Curation on the Dataverse Platform**
Dec 4, 2022 | 1:00PM - 2:00PM | Online via Zoom
Join the Harvard Dataverse Repository curation team for an insightful webinar on "Intermediate Data Curation" and maximize the potential of the Repository. This session is designed to delve into the practices, tools, and features of data curation.
- Virtual Introduction to Metadata in Dataverse Repositories: For Researchers and Support Staff**
Nov 15, 2022 | 1:30PM - 2:00PM | Online via Zoom
Come learn about how Dataverse helps researchers describe and find data. This webinar will cover how metadata fields are organized and how Dataverse features help streamline data publishing and make data more findable, accessible, and reusable.
- Virtual Introduction to Data Handling on the Dataverse Platform**
Oct 11, 2022 | 1:00PM - 2:00PM | Online via Zoom
The 2022 Dataverse Repository provides researchers with a robust data management and file handling system for sharing, managing, and preserving research datasets. The Harvard Dataverse



HARVARD UNIVERSITY
Harvard
Dataverse Support

Getting Started | Large Data Support and Services | NIH-GREI | Trainings and News | Policies & Governance | About

Go to Harvard Dataverse >

Join a growing community of Harvard and worldwide researchers who share data in the Harvard Dataverse Repository

Publishing your data and code in the Harvard Dataverse Repository:

- Increases the visibility of your research
- Gets you credit for your data through citations
- Satisfies data sharing requirements from funders and journals

Getting Started with the Harvard Dataverse Repository

Go to the Harvard Dataverse Repository

The Harvard Dataverse Repository is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard University community, where you can share, archive, cite, access, and explore research data. Each individual Dataverse (virtual repository) for organizing, managing, and showcasing datasets.

You can open your data to the general public, or restrict access and define customizable terms of use. Your data is citable and standard data citation with a Digital Object Identifier (DOI), and your metadata is open and findable.

Planning and setup

Policies & Governance

- Overview
- Terms of Use
- Privacy Policy
- Preservation Policy
- API Terms of Use
- Sample DUA

NIH-GREI

- Overview
- NIH-GREI Initiative
- Harvard Dataverse NIH-DMP Guidance
- Generalist Repository Comparison Chart

Sustainability

Data Repository Management and Curation Additional Considerations

- What features/external tool integrations will our users require?
- What file formats and sizes should be they likely to deposit?
- Are there sufficient metadata standards available to describe expected datasets?
- How will we best communicate what and who we support?
- Sensitive data considerations
- Storage limitations

Inventory of External Tools

Tool	Type	Scope	Description
Data Explorer	explore	file	A GUI which lists the variables in a table and cross tabulation analysis. https://github.com/scholarsportal/dataexplorer on adding Data Explorer
Whole Tale	explore	dataset	A platform for the creation of reproducible launch containerized interactive analysis such as Jupyter and RStudio. Using the Jupyter and RStudio environments to information, see the
Binder	explore	dataset	Binder allows you to spin up custom (including Jupyter notebooks) with https://github.com/IQSS/dataverse-binder-redirect for installation instructions.
File Previewers	explore	file	A set of tools that display the content of files - including audio, html, Hypothes.is annotations, images, GeoJSON, zip, and No the file. The previewers is using the Dataverse including how to opt through github are in Qualitative Data Rep Social Science https://github.com/scholarsportal/dataexplorer
Data Curation Tool	configure	file	A GUI for curating de assist wit https://github.com/scholarsportal/dataexplorer
Ask the Data	query	file	Ask the Data is an questions about the README.md file at https://github.com/scholarsportal/dataexplorer for the instructions
TurboCurator by ICPSR	configure	dataset	TurboCurator gene keywords. It relies TurboCurator Datavers adding

Large Data Support

Harvard Dataverse Repository now supports the sharing and storage of large data >1TB. This project is in process and content will be added as they are developed. Contact support@dataverse.harvard.edu to request large data support.

- [Globus/NESE Download Instructions \(Downloadable PDF\)](#).

Tell me what you want to know

Answer please

The dataframe has columns labeled as '1', '2', '3', '4', '5', and '6'. From the column names alone, it's not immediately clear what the data represents. To get a better understanding, we can look at a larger sample of the data or examine the values in each column. Would you like to see a larger sample of the data or explore the values in a specific column to understand the data better?

	1	2	3	4	5	6
0		0.4565123102	0.1365155458	0.2477169794	0.09832597	0.0609291946
0		0.5691616014	0.0824986391	0.2186348056	0.1089328049	0.020772149
0		0.4892006661	0.1206183581	0.2538695877	0.1055979353	0.0307134529
0		0.4788487599	0.0948798929	0.2892863839	0.1034724389	0.0335125243
0		0.5238591788	0.0813445007	0.2711584735	0.1079778489	0.015659998
0		0.4646061597	0.0860290205	0.3057681747	0.0973625504	0.0462340946
0		0.4685816472	0.1113598422	0.2900132355	0.1138951238	0.0161501512
0		0.4775180995	0.1031150551	0.279086476	0.1016715067	0.0386088627
0		0.4786808646	0.0895115773	0.2947942447	0.1192836291	0.0177296842
0		0.5306557942	0.0843725305	0.2521440939	0.09192078	0.0409068014

Viewing rows 1 through 10 of 50

Generalist Repository Ecosystem Initiative (GREI)

NIH Generalist Repository Ecosystem Initiative (GREI)

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

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

Goals of the Generalist Repository Ecosystem Initiative



1

Make it easier for researchers to **share data**.



2

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



3

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



4

Avoid duplication of the data across repositories.



5

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

NIH GREI (2022-ongoing)

NIH GREI (Generalist Repository Ecosystem Initiative)

- Current programs and activities involved:
 - Remote Large Storage Support
 - Controlled Vocabularies for Biomedical
 - Discovery for Metadata (e.g. Data Documentation Initiative (DDI))
 - Computational Workflows
 - Harvesting and Sharing Metadata Across Repositories
 - Usage Metrics - Make Data Count Support
 - Improvements for Sensitive Data Support
 - Evaluation and Evolution of Architecture
 - NIH Data Management Plans
 - Community Repository Training
 - Support for Research Objects (beyond datasets)

Participating Repositories



NIH Desirable Characteristics of Data Repositories

For Federally Funded Research

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

Guidance set forth by NIH
and by The National Science
and Technology Council, cited
in OSTP guidance





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, MeSH)
- Investigate CEDAR integration



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
 - New APIs
 - Interoperability
 - Metadata Harvesting
 - Curation Services
 - [Training and Outreach](#)

Collections/Datasets

Collection-level

Current Features



Dataverse Collections

- Own administration
- Own branding (and can be embedded anywhere)
- Metadata customization
- Search Facets

Datasets

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



Files

- Citation
- Ingest
- Preview/Explore
- Metadata
- Versioning
- Permissions/Embargo/Restrictio ns



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...

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

Other options

-
-
-
-

[Sign up for a Dataverse account.](#)



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

- General Information
- Theme + Widgets
- Permissions
- Groups
- Dataset Templates
- Dataset Guestbooks
- Featured Dataverses



DATASETS

Brain Genomics Superstruct Project (GSP) Dataverse Home
(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...

- Datasets (0)
- Datasets (1)
- Files (15)

1 to 1 of 1 Result

Publication Year
2014 (1)

Subject
Other (1)

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

Author Affiliation
Harvard University (1)

Brain Genomics Superstruct Project (GSP)
May 15, 2015

Buckner, Randy L.; Ruffman, Joshua L.; Smoller, Jordan W., 2014, "Brain Genomics Superstruct Project (GSP)", <https://doi.org/10.7910/DVN/Z5833>, 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 personal...

Collection-level

Current Features



Dataverse Collections

- Own administration
- Own branding (and can be embedded anywhere)
- Metadata customization
- Search Facets

Datasets

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



Files

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



Edit Access

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

When a user adds a new dataset to this dataverse, which role should be automatically assigned to them on that dataset?

- 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

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

Dataset-level

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
- Curation Tools (Turbo Curator)



Files

- Citation
- Ingest
- Preview/Explore
- Metadata
- Versioning
- Permissions/Embargo/Restrictio ns



Harvard Dataverse interface showing the Brain Genomics Superstruct Project (GSP) dataset page. The page includes the Harvard logo, project title, version (10.3), a citation snippet, and a description of the large-scale imaging data sets. A callout box highlights the 'Dataset Metrics' showing 13,076 Downloads.

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.

Two screenshots from the Dataverse interface. The top one shows a file list with columns for Files, Metadata, Terms, and Versions, and a search bar. The bottom one shows the 'Cite Dataset' section with the citation text: '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'.

A screenshot of the Dataverse metadata form. It includes fields for Author (Barbosa, Sonia), Affiliation (Harvard University), Identifier Scheme (ORCID, ISNI, LONA, VIAF, GND, DAI, ResearcherID, ScopusID), and another Affiliation field (Harvard University).

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**.

File-level

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
- Curation Tools



Files

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



Demographic Credibility Revised

File Name	Version	Size	Date	Access	Embargo	Restrict	Tags	Embargo	Restrict	Tags
1	1	1.0 MB	2020-08-24	Public						
2	2	1.0 MB	2020-08-24	Public						
3	3	1.0 MB	2020-08-24	Public						
4	4	1.0 MB	2020-08-24	Public						
5	5	1.0 MB	2020-08-24	Public						

Asset Citation

Specific: Almesal, Song, Lucy, 2020, "Demographic Credibility Revised", <https://doi.org/10.7910/D1/UNF-EM-JKLTJgXUJkM4KwH9w==> [9kUNF]

File Dataset - Learn about Data Citation Standards.

Preview Metadata Versions

Open in New Window

statero	staterob	year	total	land	sea	version	
1	2	USA	1816	0	0	3.2	
2	2	USA	1817	0	0	3.2	
3	2	USA	1818	0	0	3.2	
4	2	USA	1819	0	0	3.2	
5	2	USA	1820	0	0	3.2	
6	2	USA	1821	0	0	3.2	
7	2	USA	1822	0	0	3.2	
8	2	USA	1823	0	0	3.2	
9	2	USA	1824	0	0	3.2	
10	2	USA	1825	0	0	3.2	
11	2	USA	1826	0	0	3.2	
12	2	USA	1827	0	0	3.2	
13	2	USA	1828	0	0	3.2	
14	2	USA	1829	0	0	3.2	
15	2	USA	1830	0	0	3.2	
16	2	USA	1831	1	1	0	3.2
17	2	USA	1832	1	1	0	3.2
18	2	USA	1833	1	1	0	3.2
19	2	USA	1834	1	1	0	3.2
20	2	USA	1835	1	1	0	3.2
21	2	USA	1836	1	1	0	3.2
22	2	USA	1837	1	1	0	3.2
23	2	USA	1838	1	1	0	3.2
24	2	USA	1839	1	1	0	3.2
25	2	USA	1840	1	1	0	3.2
26	2	USA	1841	1	1	0	3.2
27	2	USA	1842	1	1	0	3.2

Dataverse Community

The State of the Dataverse Community is strong!

Dataverse Community (users, developers, researchers, librarians, data scientists and more)

- 120 installations across the world
- [38 countries](#)
- 1000+ subscribers to the [Google Group](#)
- 200+ people in chat.dataverse.org (Zulip)
- 1600+ GitHub issues opened by the community
- 1300+ pull requests made by the community
- 197 Github contributors to the main code base
- 186 videos on [DataverseTV](#)
- 189 community calls since 2016
- 10 years of annual [community meetings](#)
- 7 active [working groups](#)

Contributor Guide

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

New! First shipped in July 2024.

Contents:

- Ideas and Feature Requests
- Bug Reports
- Documentation
- Translations
- Code
- Usability Testing
- Answering Questions
- Sample Data
- Issue Triage
- Giving Talks
- Working Groups
- GDCC
- Artwork
- Other Contributions

Working Groups

<https://www.gdcc.io/working-groups.html>

- Containerization
- Dataverse Sustainability
- Documentation
- Frontend Development
- Large Data Support
- pyDataverse
- Sensitive Data
- UX

Google Groups

<https://dataverse.org/mailling-lists>

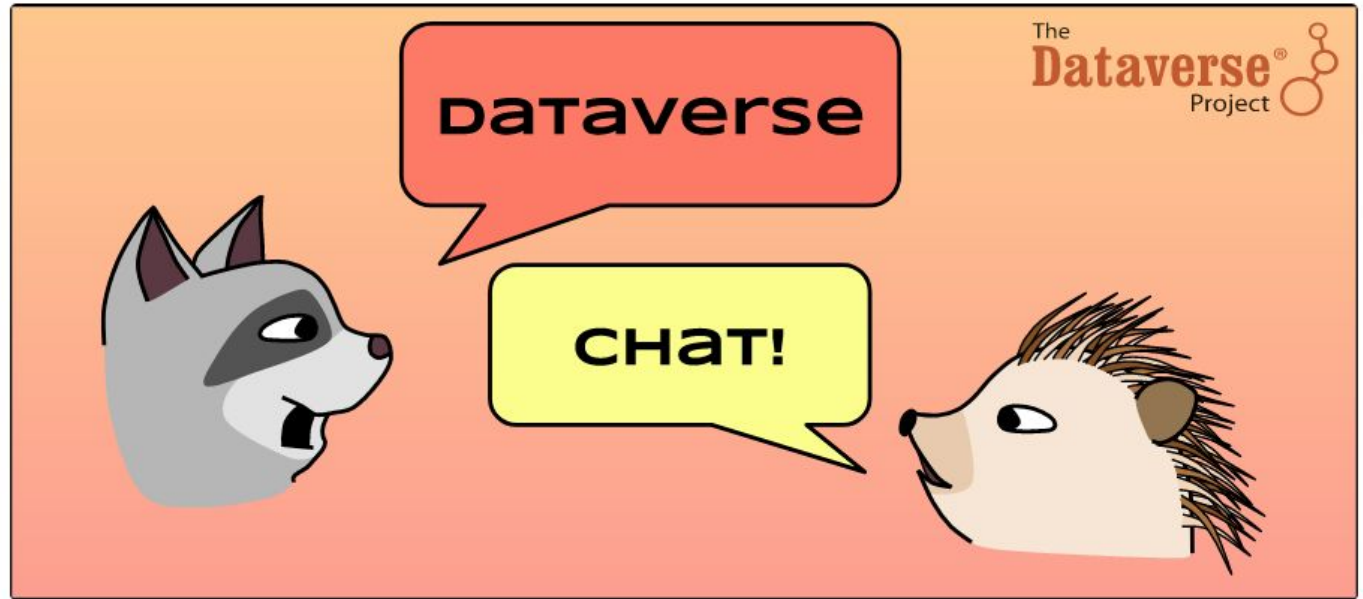
- [dataverse-community](#): for everyone!
 - most active by far
 - announcements, etc.
- [dataverse-dev](#): for developers
- [dataverse-big-data](#): for big data
- [dataverse-internationalization-wg](#): for translators
- [interessengruppe-dataverse](#): a German-language mailing list

Zulip (chat)

<https://chat.dataverse.org> or <https://dataverse.zulipchat.com>

CHANNELS

- community
- community-i18n-de
- containers
- ddi-cdi
- dev
- docs
- geospatial
- large-data
- python
- security
- troubleshooting
- ui-dev
- ux-wg
- zulip



GDCC

<https://www.gdcc.io>

DataCite membership



[About the GDCC](#) [Members](#) [Working Groups](#) [Services](#) [Sign-Up](#) [News](#)

Search



The Global Dataverse Community Consortium

Supporting Dataverse Repositories Around the World

The Global Dataverse Community Consortium (GDCC) is dedicated to providing international organization to existing [Dataverse](#) community efforts, and will provide a collaborative venue for institutions to leverage economies of scale in support of Dataverse repositories around the world.

Project board

<https://github.com/orgs/IQSS/projects/34>

The screenshot shows the GitHub Project Board for the IQSS Dataverse Project. The board is organized into three columns: "This Sprint", "In Progress", and "Ready for Review". Each card represents a project item with details such as the issue number, description, sprint, and status.

- This Sprint (10 items):**
 - dataverse-frontend #461: Collection Page: Publish Collection. Sprints: FY25 Sprint 6, GREI Re-arch. Status: Waiting.
 - dataverse-frontend #126: Add error handling mechanism to the UI. Sprints: FY25 Sprint 6, pm.GREI-d-2.71, pm.GREI-d-2.72. Status: UI.
 - dataverse-frontend #447: Add Parent Hierarchy link to Cards. Sprints: FY25 Sprint 6, GREI Re-arch. Status: SPA: Collection Page.
 - dataverse-frontend #485: Create Dataset URL. Sprints: FY25 Sprint 6, GREI Re-arch. Status: SPA: Create Dataset Form.
- In Progress (12 items):**
 - dataverse.harvard.edu #267: Implement Dropdown Selector for NIH Controlled Vocabulary in Keywords. Sprints: FY25 Sprint 4, FY25 Sprint 5, FY25 Sprint 6. Status: NIH CAFE, Size: 10, Type: Feature.
 - dataverse #10817: Fix Dataverse install scripts to handle Payara start and stop failures. Sprints: FY25 Sprint 5, FY25 Sprint 6. Status: Needs Input, Type: Bug, Size: 3.
 - dataverse #10743: Implement Dropdown for Dataset Tagging in Metadata. Sprints: FY25 Sprint 5, FY25 Sprint 6, NIH CAFE. Status: Size: 3, Type: Feature.
 - dataverse #10615: Daticite xml improvements.
- Ready for Review (6 items):**
 - dataverse #10836: 10734 Refactoring and optimization of importing and reindexing of harvested content. Feature: Harvesting. Sprints: FY25 Sprint 6. Status: Size: 30.
 - dataverse #10632: DaticiteXML changes Plus RelationType field. Consider For Next Release. Sprints: FY25 Sprint 5, FY25 Sprint 6. Status: GDCC: QDR, Size: 10.
 - dataverse-frontend #483: Private Routes Guard. Sprints: FY25 Sprint 5, GREI Re-arch. Status: SPA: Permissions, Size: 3.
 - dataverse #10829: reworked controlled vocab language keys. Sprints: FY25 Sprint 5. Status: Size: 30.

The screenshot shows the GitHub Projects page for the IQSS Dataverse Project. The page displays the project name, a search bar, and a list of projects. The project is currently open, showing a list of items with their status and sprint information.

- Project Summary:** IQSS Dataverse Project. GitHub project tracking the Dataverse Project, including Dataverse core code base and related projects and operational work for. #34 updated 3 minutes ago.
- Project Board:**
 - SPRINT- NEEDS SIZING (6 items):**
 - dataverse #10618: Tags for application container images. Champion: pdurbin. Component: Containers. Status: Needs Input, Type: Feature.
 - dataverse #10478: Tag version for the container base image. Champion: pdurbin. Component: Containers. Type: Feature. #10827.
 - SPRINT READY (38 items):**
 - dataverse #10519: Implement datasetType metadata block support (at global level). Sprints: GREI 6, GREI Year 3. Status: Size: 30, Type: Feature.
 - dataverse #6798: Feature Request/Idea: Make sure that for datasets published with CC0 waiver or standard license, metadata exports include the waiver or license. Features: Metadata, Terms & Licensing. Sprints: GREI 2. Status: Waiting for Related Issues/PRs, UX/UI Input Needed.
 - This Sprint (9 items):**
 - dataverse-frontend #461: Collection Page: Publish Collection. Sprints: FY25 Sprint 6, GREI Re-arch. Status: Size: 10, SPA: Collection Page, Waiting.
 - dataverse-frontend #126: Add error handling mechanism to the UI. Feature Request/Idea: Make sure that for datasets published with CC0 waiver or standard license, metadata exports include the waiver or license. Sprints: FY25 Sprint 6, pm.GREI-d-2.71, pm.GREI-d-2.72. Status: Size: 30, UI.
 - dataverse-frontend #447: Add Parent Hierarchy link to Cards. Sprints: FY25 Sprint 6, GREI Re-arch. Status: SPA: Collection Page, Size: 3.
 - dataverse-frontend #485: Create Dataset URL. Sprints: FY25 Sprint 6, GREI Re-arch. Status: SPA: Create Dataset Form, Size: 10.

From bug discovery to fix (main code base)

<https://guides.dataverse.org/en/6.3/developers/version-control.html#how-to-make-a-pull-request>

- How to Make a Pull Request
 - Find or Create a GitHub Issue
 - Finding GitHub Issues to Work On
 - Creating GitHub Issues to Work On
 - Communicate Which Issue You Are Working On
 - Create a New Branch Off the develop Branch
 - Commit Your Change to Your New Branch
 - Writing a Release Note Snippet
 - Push Your Branch to GitHub
 - Make a Pull Request
 - Make Sure Your Pull Request Has Been Advanced to Code Review

Adding features (main code base)

- Please reach out before coding.
- Design docs are great.
 - e.g. [Proposal: Improving dataset relationships in Dataverse](#)

Creating GitHub Issues to Work On

You are very welcome to create a GitHub issue to work on. However, for significant changes, please reach out (see [Getting Help](#)) to make sure the team and community agree with the proposed change.

For small changes and especially typo fixes, please don't worry about reaching out first.

Contributing to the new frontend (React SPA)

<https://github.com/IQSS/dataverse-frontend#contributing>

<https://ui.gdcc.io>

Contributing

We love PRs! Read the [Contributor Guidelines](#) for more info. Any contributions you make are **greatly appreciated**.

Got Questions? Join the conversation on [Zulip](#), or our Google Groups for [Developers](#) and [Users](#). Or attend community meetings, hosted by the Global Dataverse Community Consortium to collaborate with the interest groups for [Frontend Development](#) and [Containerization](#), learn and share with communities around the world!

Contributing to other code bases

<https://guides.dataverse.org/en/6.3/contributor/code.html>

Many Codebases, Many Languages

The primary codebase and issue tracker for Dataverse is <https://github.com/IQSS/dataverse>. It's mostly backend code written in Java. However, there are many other codebases you can work on in a variety of languages. Here are a few that are especially active:

- <https://github.com/IQSS/dataverse> (Java)
- <https://github.com/IQSS/dataverse-frontend> (React)
- <https://github.com/IQSS/dataverse-client-javascript> (TypeScript)
- <https://github.com/gdcc/dataverse-previewers> (Javascript)
- <https://github.com/gdcc/pyDataverse> (Python)
- <https://github.com/gdcc/dataverse-ansible> (Ansible)
- <https://github.com/gdcc/dv-metrics> (Javascript)

If nothing above sparks joy, you can find more projects to work on under [Client Libraries](#), [Building External Tools](#), [Related Projects](#), and [Apps](#).

Getting in touch

- Zulip - <https://chat.dataverse.org>
- Tech hours on Tuesdays at 3pm Eastern Time
- Community calls - <https://dataverse.org/community-calls>
- Dataverse Community Meeting: <https://dataverse.org/events>