

# The Use Experience in a Dataverse Supported Repository


Sonia Barbosa

National Yang Ming Chiao Tung University, February 2024


# Agenda

## AGENDA: The User Experience

- Community
- Collections
- Datasets
- Files
- Files and External Tools
- Permissions and Access
- Data Exploration
- Search and Browse Functionality


 Dataverse Project

About▼Community▼Best Practices▼Software▼Contact




### The Dataverse<sup>®</sup> 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 norms for sharing, preserving, citing, exploring, and analyzing research data. Contribute code extensions, documentation, testing, and/or standards. *Integrate research analysis, visualization and exploration tools*, or other research and data archival systems with the Dataverse Project. [Want to contribute?](#)

# 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>

# The Dataverse Project and FAIR



# FAIR Principles and Dataverse

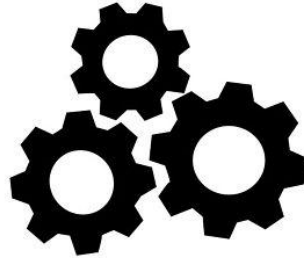
F<sub>indable</sub>



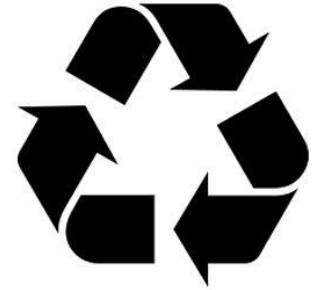
A<sub>ccessible</sub>



I<sub>nteroperable</sub>



R<sub>eusable</sub>



[https://en.wikipedia.org/wiki/File:FAIR\\_data\\_principles.jpg](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 (meta)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

# The User Experience - an Overview

IQSS / dataverse

Code

Issues 1.1k

Pull requests 106

Actions

Projects

Security 26

Insights

Settings

dataverse

Public

develop 256 Branches67 Tags

Go to file

pdurbin Merge pull request #10250 from IQSS/10249\_search\_api\_type

github

Add default label to request issue to

conf

Merge branch 'develop' into 6783-c

doc

Merge pull request #10250 from IQSS

modules

Bump version to 6.1

scripts

Add configuration for automatic XH

src

Merge pull request #10269 from Qu

tests

add S3AccessIT to the list of integr

env

update Solr 8.11.1 to 9.3.0 (minus c

.gitattributes

Add the .gitattributes file to set LF

.gitignore

remove Vagrant (unused) #9838

.readthedocs.yml

ci(docs): make RTFD install graphviz, too

Groups

Conversations

Search conversations within dataverse-com...

1-30 of 2528

Dataverse Users Community 982 members

Welcome to the Dataverse Users Community Group! Please feel free to ask a question, share feedback or start a discussion about anything and everything involving Dataverse. For more information about the Dataverse Project please visit: <https://dataverse.org>

New conversation

My groups

Recent groups

All groups

Local Contexts

Public groups

Starred conversations

Dataverse Users Community

Conversations 99+

Approved 99+

Pending

Labels

People

Members

Pending members

Banned users

About

My membership settings

Group settings

Report repository

Releases 66

5 months ago

8 months ago

docker-compose-dev.yml

Update docker-compose-dev.yml

4 days ago

pom.xml

Merge remote-tracking branch 'origin/develop' into 9920...

last month

README

Code of conduct

License

Security

Dataverse®

Dataverse is an [open source](#) software platform for sharing, finding, citing, and preserving research data (developed by the [Dataverse team](#) at the [Institute for Quantitative Social Science](#) and the [Dataverse community](#)).  
[dataverse.org](#) is our home on the web and shows a map of Dataverse installations around the world, a list of [features](#), [integrations](#) that have been made possible through [REST APIs](#), our development [roadmap](#), and more.  
We maintain a demo site at [demo.dataverse.org](#) which you are welcome to use for testing and evaluating Dataverse.  
To install Dataverse, please see our [Installation Guide](#) which will prompt you to download our [latest release](#).  
To discuss Dataverse with the community, please join our [mailing list](#), participate in a [community call](#), chat with us at [chat.dataverse.org](#), or attend our annual [Dataverse Community Meeting](#).  
We love contributors! Please see our [Contributing Guide](#) for ways you can help.  
Dataverse is a trademark of President and Fellows of Harvard College and is registered in the United States.

The Dataverse Project

API Test Status

API Test Coverage inaccessible

Maven Tests passing

Unit Test Coverage 80%

Guides Build Status passing

+ 170 contributors

Deployments 32

beta-testing

+ 31 deployments

Languages

Java 82.6%

HTML 10.2%

JavaScript 3.3%

Shell 1.3%

Python 0.8%

XSLT 0.6%

Other 1.2%

For you

Following

Home

Explore

Notifications

Messages

Grok

Lists

Bookmarks

Communities

Premium

Profile

More

Post

What is happening?!

Post

Slava Tykhonov @4tykhonov · Jan 27 · 🌐

My latest experiments on the Open Source [LLM](#) models' 'decompression' show great potential in transforming all "packaged" content into a structured format and ingesting it back into the knowledge graph. Even in the current state of AI, Generative AI is capable of saving...  
[Show more](#)

Skosmos

Queen Island thesaurus

Made in Heaven

1 3 5 612

Slava Tykhonov @4tykhonov · 1h

It looks like this can be considered as official start of the battle between human and [AI](#) in the field of information classification and ontologies. And we're in the very beginning...

RRE

Infectious disease

Infectious disease

1 59

OpenCitations @opencitations Follow

CORE @iucore Follow

Melissa Sands @melissaleess Follows you Follow

Show more

Terms of Service Privacy Policy Cookie Policy Accessibility Ads info More ... © 2024 X Corp.

# ➡ 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



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.](#)

# Collection-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



### Files

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



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... Advanced Search Add Data

Dataverses (0)  
Datasets (1)  
Files (15)

Publication Year  
2014 (1)

Subject  
Other (1)

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

Author Affiliation  
Harvard University (1)

1 to 1 of 1 Result

Brain Genomics Superstruct Project (GSP)  
May 15, 2015

Buckner, Randy L.; Roffman, Joshua L.; Smoller, Jordan W., 2014, "Brain Genomics Superstruct Project (GSP)",  
<https://doi.org/10.7910/DVN/25853>, 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...

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

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

# 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



### Files

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

**HARVARD**  
Dataverse

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

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

**Brain Genomics Superstruct Project (GSP)**  
Version 1.0.5

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

**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.

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

**Subject**  
Other  
License/Data Use Agreement Custom Dataset Terms

**Files**  
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  
GSP\_TSP\_2017\_105  
GSP\_TSP\_2017\_105  
GSP\_TSP\_2017\_105  
GSP\_TSP\_2017\_105  
GSP\_TSP\_2017\_105  
GSP\_TSP\_2017\_105  
GSP\_TSP\_2017\_105  
GSP\_TSP\_2017\_105  
GSP\_TSP\_2017\_105

**Cite Dataset** Learn about **Data Citation Standards**.

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

# 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



### Files

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

Subject: 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

1 GSP\_DataUse\_Terms\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

2 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

3 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

4 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

5 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

6 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

7 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

8 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

9 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

10 GSP\_140422.pdf  
Author: GSP, 2014  
Published: Aug 24, 2014  
1,130 Downloads

Edit Files

- Metadata
- Restrict
- Unrestrict
- Tags
- Embargo
- Delete

Demographic Credibility Revised

Specific: Aleson, Song, Lucy, 2020, "Demographic Credibility Revised", <https://doi.org/10.7910/D11UNF-6mJkL8yXUkM4K6l9w> [Full UNF]

File Dataset - Learn about Data Citation Standards.

Preview Metadata Versions

Open in New Window

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

File Tools Metadata Versions

File Tools - Open in New Window

Tell me what you want to know

What is this data about?

Answer please

The data table contains information on various aspects of government and society, such as foreign relations, planning and development, construction and public works, finance, budget and treasury, government, interior and home affairs, health and social welfare, environment, communications and information, transport, and agriculture, food, fisheries and livestock.

cid	iso3	year	leadcode	portfolio_1	IMFn	IMFnset	IMFongoing	maleshare_minister	male	female	mingr	genderpost	prestpost	incid	f
3	AFG	1980	170	Foreign Relations	0	0	0	1	1	0	0	-1	1	2	0
3	AFG	1980	170	Planning & Development	0	0	0	1	1	0	0	0	0	2	0
3	AFG	1980	170	Construction & Public	0	0	0	1	1	0	0	-1	0	2	0

Asset Citation

Specific: Aleson, Song, Lucy, 2020, "Demographic Credibility Revised", <https://doi.org/10.7910/D11UNF-6mJkL8yXUkM4K6l9w> [Full UNF]

File Dataset - Learn about Data Citation Standards.

Preview Metadata Versions

Open in New Window

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



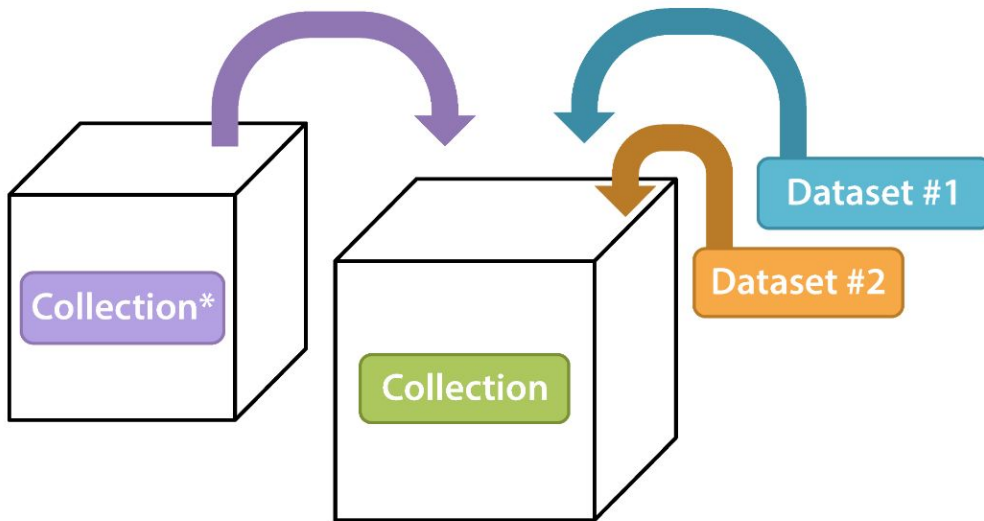
# The User Experience: Collections

# Create a 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

- 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

\*contact [support@dataverse.harvard.edu](mailto:support@dataverse.harvard.edu) and request the creation of a collection




[DANIEL P. ALDRICH](#)[BIO & C.V.](#)[RESEARCH](#)[DATA](#)[PUBLICATIONS](#)[MEDIA](#)[more...](#)




## DANIEL P. ALDRICH

Daniel P. Aldrich is full professor of political science and Director of the [Security and Resilience Studies Program](#) at [Northeastern University](#). He researches post-disaster recovery, countering violent extremism, the siting of controversial facilities and the interaction between civil society and the state. He has held posts as a Fulbright Research Fellow and an Abe Fellow at Tokyo University and as an AAAS Science and Technology Fellow with USAID. Aldrich is a contributor to the New York Times, CNN, The Conversation, and the Asahi Shimbun, among other media.

**HARVARD**  
Dataverse


[Add Data](#) [Search](#) [About](#) [User Guide](#) [Support](#) [Sign Up](#) [Log In](#)





[Daniel Aldrich Dataverse](#) [Northeastern University](#)  
(Northeastern University)


[Harvard Dataverse >](#)

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

 [Advanced Search](#)


☒  [Dataverses \(0\)](#)

☒  [Datasets \(10\)](#)


☐  [Files \(27\)](#)

**Publication Year**  
2022 (1)  
2020 (3)  
2019 (2)  
2016 (1)  
2012 (1)  
[More...](#)

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



**Replication Data for: How Social Infrastructure Saves Lives: A Quantitative Analysis of Japan's 3/11 Disasters**  
Nov 23, 2022  
Aldrich, Daniel, 2022, "Replication Data for: How Social Infrastructure Saves Lives: A Quantitative Analysis of Japan's 3/11 Disasters", <https://doi.org/10.7910/DVN/YA0EYS>, Harvard Dataverse, V1, UNF:6:fe9B3jEgPRcE25dCykAt6g== [fileUNF]  
Observers have long debated how societies should invest resources to safeguard citizens and property, especially in the face of increasing shocks and crises. This article explores how social infrastructure -the spaces and places that help build and maintain social ties and trust,...




**Brewing Social Capital: A Survey of Breweries and Social Ties across the United States**  
Dec 20, 2020  
Aldrich, Daniel; Schulz, Eric, 2020, "Brewing Social Capital: A Survey of Breweries and Social Ties across the United States",

陽明交大  
NYCU

Add Data ▾ Search ▾ User Guide

NYCU Dataverse

Metrics 1,741 Downloads


  
Wei-Chen (Walon) Chiu Dataverse

Search this dataverse...

Advanced Search

☒ Datasets (1)


☒ Datasets (177)

☐ Files (4,802)

**Dataverse Category**  
 Researcher (1)

**Publication Year**

1 to 10 of 178 Results


  
DenseNet-Powered Radiologist Early Warning System by RGB Imaging  
 Jan 31, 2024  
 Shyh-Wei Chen; Yu-Heng Hsieh; Jyun-Kai Chen; Shin-Fu Wang; Wei-Ching-Tsornng Tsai; Jyh-Wen Chai; Shyan-Ming Yuan, 2024, "DenseNet-Powered Radiologist Early Warning System by RGB Imaging", NYCU Dataverse, V1  
<https://doi.org/10.57770/12YL4C>, NYCU Dataverse, V1

This is a research paper of radiologist early warning system using densenet

**Author Name**

Wei-Chen Chiu (21)  
 Tien, Chih Shan (21)  
 Shang-Hsiung Yang (19)  
 Yau-Huei Wei (15)  
 Y. H. Wei (14)  
 Yi-Hsuan Tsai (13)  
 Min Sun (7)  
 Shyan-Ming Yuan (7)  
 Yu-Ting Wu (7)  
 Yuan, Shyan-Ming (6)  
 Fu-En Wang (4)  
 Jin-Wu Tsai (4)  
 Mario Fritz (4)  
 Y. T. Wu (4)  
 Yi-Ju Tseng (4)  
 Less... More...

陽明交大  
NYCU

Add Data ▾ Search ▾ User Guide ▾ About ▾ Support ▾ English ▾ Log In


  
Wei-Chen (Walon) Chiu Dataverse Visit my personal website  
 (Computer Science, NYCU)

NYCU Dataverse >

Associate Professor in Computer Science at National Chiao Tung University  
 Frank S Chen Junior Chair Professor cacaFly Junior Chair Professor  
 Doctor of Engineering Science (Dr.-Ing.) Max Planck Institute for Informatics

Search this dataverse...

Advanced Search

☒ Datasets (2)

☒ Datasets (3)

☐ Files (1,074)

**Publication Year**  
 2022 (3)

**Subject**  
 Computer and Information Science (3)

**Author Affiliation**  
 Max Planck Institute for Informatics (2)  
 Department of Computer Science, National Chiao Tung University (1)

1 to 3 of 3 Results

IT Sort ▾

Replication Data for: Using generification to create and label photos that are challenging for computer vision and people  
 Jun 14, 2022  
 Peter Kolinski; Xi-Jing Zhang; Yeng Chieh-Yun; Wei-Chen Chiu; Yung-Ju Chang, 2022, "Replication Data for Using generification to create and label photos that are challenging for computer vision and people", <https://doi.org/10.57770/12YL4C>, NYCU Dataverse, V1  
 It would be hard to overstate the importance of Computer Vision (CV), applications of which can be found from self-driving cars, through facial recognition to augmented reality and the healthcare industry. Recent years have witnessed dramatic progress in visual-object recognition...

Replication Data for: Improving the Kinect by Cross-Modal Stereo  
 Jun 14, 2022  
 Wei-Chen Chiu; Ulf Bleske; Mario Fritz, 2022, "Replication Data for Improving the Kinect by Cross-Modal Stereo", <https://doi.org/10.57770/12YL4C>, NYCU Dataverse, V1  
 Kinect's active sensing strategy is very well suited to produce robust and high-frame rate depth maps for human pose estimation. But the shift to the industrial domain surfaced applications under a wider set of operation conditions it wasn't originally designed for. We use the sensor...

Replication Data for: Multi-Class Video Co-Segmentation with a Generative Multi-Video Model  
 Jun 14, 2022  
 Wei-Chen Chiu; Mario Fritz, 2022, "Replication Data for Multi-Class Video Co-Segmentation with a Generative Multi-Video Model", <https://doi.org/10.57770/12YL4C>, NYCU Dataverse, V1  
 Video data provides a rich source of information that is available to us today in large quantities e.g. from online resources. Tasks like segmentation benefit greatly from the analysis of spatio-temporal motion patterns in videos and recent advances in video segmentation has shown...

Wei-Chen (Walon) Chiu; Taiwanese: Khu Uî-Sîn

Home About Me Research Contact



Associate Professor in Computer Science  
 at National Chiao Tung University  
 Frank S Chen Junior Chair Professor  
 cacaFly Junior Chair Professor  
 Doctor of Engineering Science (Dr.-Ing.)  
 Max Planck Institute for Informatics


 walon [at] cs [dot] nctu [dot] edu [dot] tw  
 walon [dot] chiu [at] gmail [dot] com

 Curriculum Vitae

 Google Scholar

[Office] Room 526, Engineering Building C, Guangfu Campus, NCTU  
 [Telephone] +886-3-5712121 ext. 54730

Information about my research group:  
 [Name] Enriched Vision Applications Lab  
 [Location] Room 548, Engineering Building C, Guangfu Campus, NCTU



# Metadata you can use to describe collections

- Names
- Affiliations
- Identifiers of collection URLs
- Categories
- Emails
- Descriptions

New Dataverse

[Demo Dataverse](#) >

\*Asterisks indicate required fields

**Host Dataverse** ?

Enter Dataverse Name

**Dataverse Name** \* ?

Julian Gautier Dataverse

**Affiliation** ?

Harvard University

**Identifier** \* ?

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

**Category** \* ?

Select one...

**Email** \* ?

juliangautier@g.harvard.edu

**Description** ?

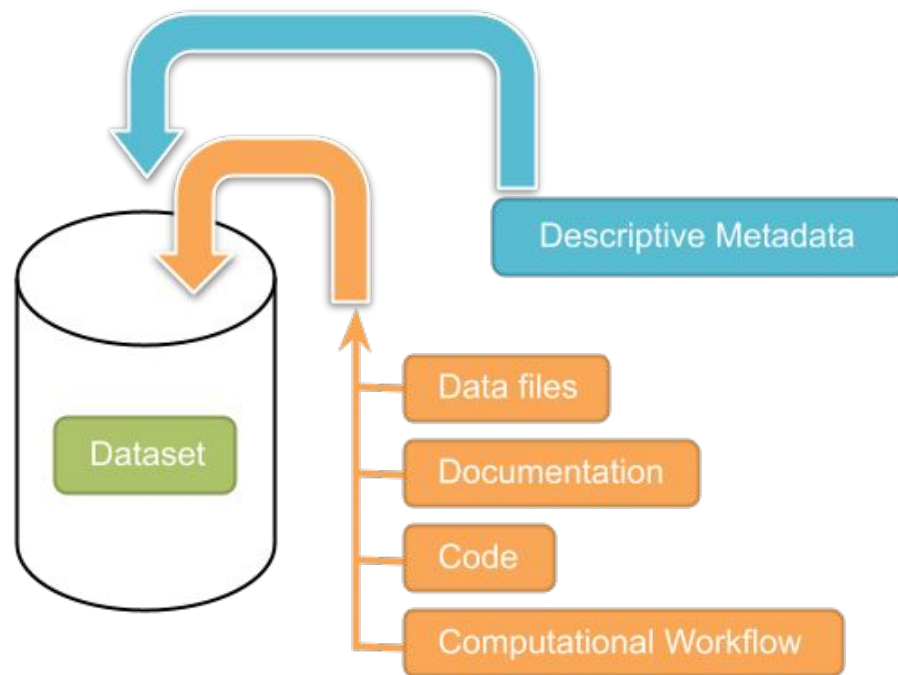
This field supports only certain [HTML tags](#).

# The User Experience: Datasets

# 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.**



## Cambodia 2020 Quick Start

(Purdue University Main Campus)

Harvard Dataserve > International Programs in Agriculture (PIA) - Purdue University > Feed the Future Innovation Lab for Food Safety > Cambodia > Cambodia 2020 Quick Start >

# Distribution Center Generic E. Coli, Coliform, and Enterobacteriaceae Detection and Quantification Data

Version 1.0



Vicham, Jessie, 2022, "Distribution Center Generic E. Coli, Coliform, and Enterobacteriaceae Detection and Quantification Data", <https://doi.org/10.7910/DVN/0XUF4S>, Harvard Dataserve, V1, UNF:6:XY1pY4yKlRyzZkR19Cn3A== [fileUNF]

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

Access Dataset -

Contact Owner Share

Dataset Metrics

12 Downloads

## Description

The dataset includes data that was collected from tomatoes, cucumbers, and lettuce from a distribution center located in Cambodia. This dataset includes detection and quantification data for all vegetables types, in two seasons (rainy and dry) for generic E. coli, Coliforms, and Enterobacteriaceae.

Agricultural Sciences

## Subject

## Keyword

Cambodia, Vegetables, Food Safety, Cambodia, vegetables, food safety, tomatoes, cucumbers, lettuce, coliform bacteria, Enterobacteriaceae

## License/Data Use Agreement



Files Metadata Terms Versions

Search this dataset...

Filter by

File Type: All - Access: All -

Sort

1 to 2 of 2 Files

Download



Cambodia\_Quantification\_codebook.tab

Tabular Data - 581 B

Published Sep 7, 2022

6 Downloads

3 Variables, 7 Observations UNF:6:atiQ...CXw== [codebook for Cambodia quantification data]



Cambodia\_Quantification\_final.tab

Tabular Data - 14.5 KB

Published Sep 7, 2022

6 Downloads

7 Variables, 384 Observations UNF:6:QIN...qQQ== [ ]



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

## Brain Genomics Superstruct Project (GSP)

Version 0.0



Buckner, Randy L., Hoffman, Joshua L., Smoller, Jordan W., 2014, "Brain Genomics Superstruct Project (GSP)", <https://doi.org/10.7910/DVN/025633>, Harvard Dataserve, V10

Cite Dataset -

Learn about [Data Citation Standards](#).

Access Dataset -

Contact Owner Share

Dataset Metrics

14,434 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.

Please provide an academic or research institution email address when applying for access.

## Subject

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 -

Sort

1 to 10 of 15 Files

Download

Request Access



GSP\_DataUse\_Terms\_140422.pdf

Adobe PDF - 282.7 KB

Published Aug 24, 2014

1,435 Downloads

MD5: e64...209 [ ]

Data use terms

Documentation



GSP\_1st\_140630.csv

Plain Text - 836.4 KB

Published Aug 24, 2014

840 Downloads

MD5: 27e...4d8 [ ]

Demographic, cognitive/behavior, quality control, and morphometrics data.

Documentation



GSP\_part10\_140630.tar

GNU Archive - 10.0 GB

Published May 21, 2014

1,182 Downloads

MD5: 302...4d8 [ ]

Tar archive of imaging data for subjects 1474-1572; refer to GSP\_README\_140630.pdf for more information.



GSP\_part11\_140630.tar

GNU Archive - 10.1 GB

Published May 21, 2014

1,209 Downloads

MD5: a62...5d9 [ ]

Tar archive of imaging data for subjects 1474-1572; refer to GSP\_README\_140630.pdf for more information.



GSP\_part12\_140630.tar

GNU Archive - 10.2 GB

Published May 21, 2014

1,204 Downloads

MD5: 165...308 [ ]

Tar archive of imaging data for subjects 158-314; refer to GSP\_README\_140630.pdf for more information.



GSP\_part13\_140630.tar

GNU Archive - 10.1 GB

Published May 21, 2014

1,163 Downloads

MD5: 17e...4d8 [ ]

Tar archive of imaging data for subjects 315-471; refer to GSP\_README\_140630.pdf for more information.



GSP\_part14\_140630.tar

GNU Archive - 9.7 GB

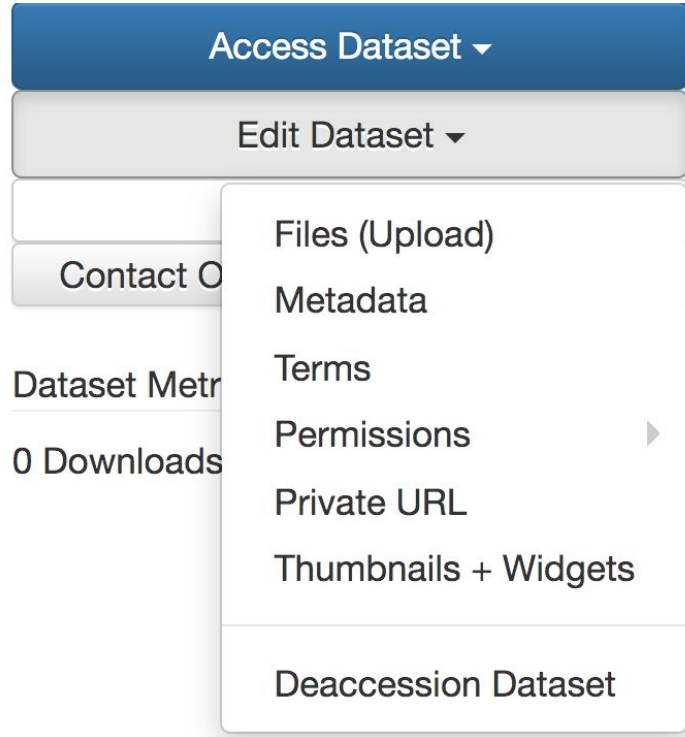
Published May 21, 2014

1,163 Downloads

MD5: 17e...4d8 [ ]

Tar archive of imaging data for subjects 315-471; refer to GSP\_README\_140630.pdf for more information.

# Datasets - Editing Options



- A **Private URL** can be generated for DRAFT datasets to share a dataset with a coauthor or journal, before publishing.

# Metadata you can use to describe datasets

Using many fields shown  
in collapsible panels,  
such as “Citation  
Metadata” and “Terms”

\*Asterisks indicate required fields

### Citation Metadata ^

<b>Title *</b> ?	<input type="text"/>		
	<button>Add "Replication Data for" to Title</button>		
<b>Author *</b> ?	<b>Name *</b> ?	<b>Affiliation</b> ?	
	<input type="text" value="Gautier (SU), Julian"/>	<input style="border: 1px solid #ccc;" type="text" value="Harvard University"/>	<input style="border: 1px solid #ccc;" type="button" value="+"/>
		<b>Identifier Type</b> ?	
		<input style="border: 1px solid #ccc;" type="text" value="Select..."/>	
	<b>Identifier</b> ?		
	<input type="text"/>		
<b>Point of Contact *</b> ?	<b>Name</b> ?	<b>Affiliation</b> ?	
	<input type="text" value="Gautier (SU), Julian"/>	<input style="border: 1px solid #ccc;" type="text" value="Harvard University"/>	<input style="border: 1px solid #ccc;" type="button" value="+"/>
	<b>E-mail *</b> ?		
	<input type="text" value="juliangautier@g.harvard.edu"/>		
<b>Description *</b> ?	This field supports only certain <a href="#">HTML tags</a> .		
	<b>Text</b> ?		
	<input type="text"/>	<input style="border: 1px solid #ccc;" type="button" value="+"/>	

# Metadata Dataverse creates for datasets

- Persistent IDs
- Creation dates
- Publication dates
- Version numbers
- Sum of file downloads
- Sum of page views  
(in repositories with Make Data Count enabled)

The screenshot displays the 'Metadata' tab of a Dataverse dataset page. At the top, there are four tabs: 'Files', 'Metadata', 'Terms', and 'Versions'. Below the tabs, there are two buttons: 'Add + Edit Metadata' (with a pencil icon) and 'Export Metadata' (with a download icon and a dropdown arrow). The main content area is titled 'Citation Metadata' with an upward arrow. It contains three rows of metadata:

<b>Persistent Identifier</b> ?	doi:10.7910/DVN/VZ5JFG
<b>Publication Date</b> ?	2023-11-13
<b>Title</b> ?	Replication Data for: "Prevalence of Neuromyths Among Students and Pre-

# Metadata Dataverse creates for datasets

- Differences among versions

	Dataset Version	Summary	Contributors	
<input type="checkbox"/>	12.0	<b>Citation Metadata:</b> Description (1 Changed); Notes (Changed); <b>Additional Citation Metadata:</b> (3 Changed); <b>Files (Added: 98; Removed: 84);</b> <a href="#">View Details</a>	Julian Gautier	202
<input type="checkbox"/>	11.0	<b>Citation Metadata:</b> Description (1 Changed); Notes (Changed); <b>Additional Citation Metadata:</b> (4 Changed); <b>Files (Added: 84; Removed: 57);</b> <a href="#">View Details</a>	Julian Gautier	202
<input type="checkbox"/>	10.0	<b>Citation Metadata:</b> Notes (Changed); Description (1 Changed); <b>Additional Citation Metadata:</b> (2 Removed, 2 Changed); <b>Files (Added: 57; Removed: 50);</b> <a href="#">View Details</a>	Julian Gautier	202
<input type="checkbox"/>	9.0	<b>Citation Metadata:</b> Description (1 Changed); <b>Additional Citation Metadata:</b> (1 Added); <b>Files (Added: 50; Removed: 37);</b> <a href="#">View Details</a>	Julian Gautier	202
<input type="checkbox"/>	8.1	<b>Citation Metadata:</b> Description (1 Changed); <a href="#">View Details</a>	Julian Gautier	202
<input type="checkbox"/>	8.0	<b>Citation Metadata:</b> Description (1 Changed); <b>Additional Citation Metadata:</b> (1 Added, 1 Changed); <b>Files (Added: 1; Replaced: 1);</b> <a href="#">View Details</a>	Julian Gautier	202
<input type="checkbox"/>	7.0	<b>Citation Metadata:</b> Description (1 Changed); <b>Files (Added: 35; Removed: 36);</b> <a href="#">View Details</a>	Julian Gautier	202

# The User Experience: Files

# File sizes and upload options and examples

## File formats:

- All formats accepted

## Per file size limitations\*:

- Set by repository

## Tabular Ingest\*:

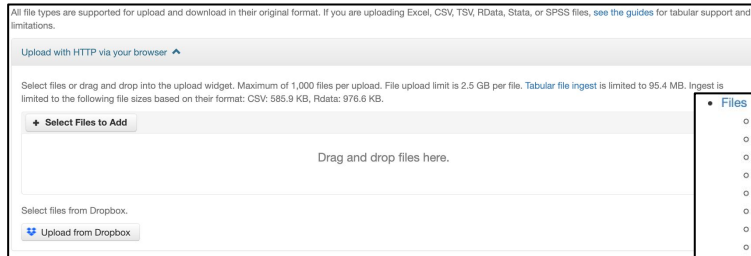
- Set by repository

## Total project size limitations\*:

- 1 TB (non Harvard)
- 2.5TB (Harvard)
- Paid options available for larger projects

## Policies\*:

- No identifiable data
- Copyright adherence
- No protected password files
- Terms required (CC0 by default)



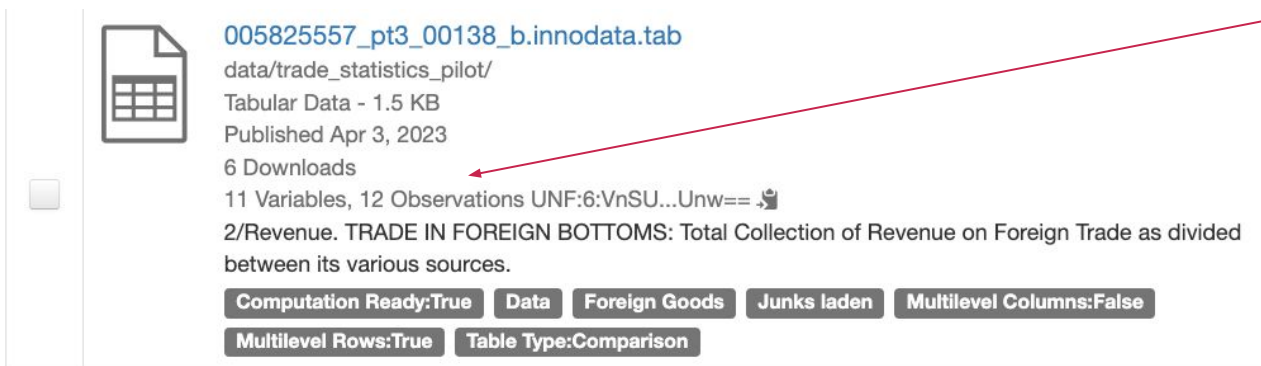
## File upload options

## Native APIs file handling options

- Files
  - Get JSON Representation of a File
  - Adding Files
  - Accessing (downloading) files
  - Restrict Files
  - Uningest a File
  - Reingest a File
  - Redetect File Type
  - Extract NcML
  - Replacing Files
  - Deleting Files
  - Getting File Metadata
  - Updating File Metadata
  - Editing Variable Level Metadata
  - Provenance
    - Get Provenance JSON for an uploaded file
    - Get Provenance Description for an uploaded file
    - Create/Update Provenance JSON and provide related entity name for an uploaded file
    - Create/Update Provenance Description for an uploaded file
    - Delete Provenance JSON for an uploaded file
  - Datafile Integrity
  - Get External Tool Parameters
  - Get Fixity Algorithm

**\*Determined by repository.** Unless otherwise stated, these policies are for Harvard repository.

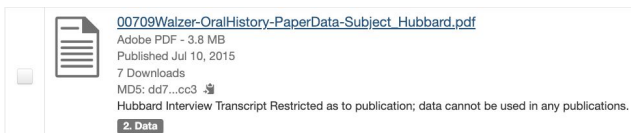
# File level metadata considerations and features



005825557\_pt3\_00138\_b.innodata.tab  
data/trade\_statistics\_pilot/  
Tabular Data - 1.5 KB  
Published Apr 3, 2023  
6 Downloads  
11 Variables, 12 Observations UNF:6:VnSU...Unw== .  
2/Revenue. TRADE IN FOREIGN BOTTOMS: Total Collection of Revenue on Foreign Trade as divided between its various sources.

Computation Ready:True Data Foreign Goods Junks laden Multilevel Columns:False  
Multilevel Rows:True Table Type:Comparison

Metadata extracted from tabular and other files with tabular ingest feature: **Observations, Variables**



00709Walzer-OralHistory-PaperData-Subject\_Hubbard.pdf  
Adobe PDF - 3.8 MB  
Published Jul 10, 2015  
7 Downloads  
MD5: dd7...cc3  
Hubbard Interview Transcript Restricted as to publication; data cannot be used in any publications.

2. Data

File level “terms”



005825557\_pt3\_00140.innodata.txt  
data/trade\_statistics\_pilot/  
Plain Text - 194 B  
Published Apr 3, 2023  
0 Downloads  
MD5: e18...5a8  
File associated with data tables series: Population:15

Customs Data JAMES H. HART STATISTICS

File level “tags”

Plan to populate additional metadata for files: description, edit file names, add tags



# Towards AI integration

**HARVARD**  
Dataverse

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

Harvard Dataverse >

## Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends

Version 1.0

Koren, Ore, 2017, "Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends", <https://doi.org/10.7910/DVN/PHHZI7>, Harvard Dataverse, V1, UNF:6NDQ2oRK6QGGA+OZgUMdA== [fileUNF]

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

**Description**

Research into the causes of violence against civilians has increased significantly in recent years, yet the mechanisms governing spatial patterns of victimization remain poorly understood. My investigation explores if and why one specific locality, capital cities, experiences a higher frequency of violence against civilians perpetrated by armed insurgent organizations. I argue that the political value associated with capitals allows these groups to asymmetrically impose higher costs on the regime by targeting civilians in these localities. I lay out and validate three specific mechanisms to explain this pattern: elite coercion, popular intimidation, and international persuasion. In the first scenario insurgents aim to influence domestic elites directly. In the second, they aim to affect domestic civilians' resolve. In the third, they seek to influence international audiences. Using new geolocated global atrocities data for the years 1996-2009, I evaluate this linkage by employing different methodological approaches and accounting for potential reporting biases. Finally, I show that ethnic and secessionist wars are more likely to experience atrocities in the capital compared with other conflicts. The findings illustrate potential benefits from explaining the temporal and spatial variation in violence by insurgents, with a focus on strategic conditions and power asymmetries. (2017-09-21)

**Subject**  
Social Sciences

**Keyword**  
Political violence; Civilian victimization; Conflict; Spatial analysis

**License/Data Use Agreement**  
 CC0 1.0

Files Metadata Terms Versions

**Change View** Table Tree

Search this dataset...

Filter by  
File Type: All - Access: All -

Sort -

21 to 30 of 43 Files

Download -

	A.8. Forecasting and Cross Validation Exercises.R Replication files 09_20_17/ R Syntax - 4.9 KB Published Sep 21, 2017 39 Downloads MDS: 9f8...7a5	
	A.9. Capital Atrocities and Regime Survival.R Replication files 09_20_17/ R Syntax - 11.1 KB Published Sep 21, 2017 39 Downloads MDS: 7c1...990	
	eth_sec_war_dat.tab Replication files 09_20_17/Data/Conflict Type Analysis/ Tabular Data - 50.5 MB Published Sep 21, 2017 46 Downloads 48 Variables, 214840 Observations UNF:6:URZ8...MBw==	

<https://doi.org/10.7910/DVN/PHHZI7>



This chatbot only sees the tabular data but is clueless about the metadata

# Tell me what this data are about

[Harvard Dataverse](#) > [Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends](#) >

## afginc.tab

This file is part of "Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends".

Version 1.0

### File Citation

Koren, Ore, 2017, "afginc.tab", *Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends*, <https://doi.org/10.7910/DVN/PHHZI7/GVJQH3>, Harvard Dataverse, V1, UNF:6:nUddO4V8XEZZCnhi+9rxYg== [fileUNF]

[Cite Data File](#)

Learn about [Data Citation Standards](#).

### Dataset Citation

Koren, Ore, 2017, "Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends", <https://doi.org/10.7910/DVN/PHHZI7>, Harvard Dataverse, V1, UNF:6:NDQ2oRK6QGGrA+OZgUMdDA== [fileUNF]

[Cite Dataset](#)

Learn about [Data Citation Standards](#).

**Access File**

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

File Metrics

50 Downloads

[File Tools](#) [Metadata](#) [Versions](#)

[File Tools](#) [Open in New Window](#)

Tell me what you want to know

What is this data about?

Answer please

This data is about latitude, longitude, and event information.


Unnamed: 0	latitude	longitude	event
1	30.9666666667	61.8833333333	2
2	31	64	3
3	31.0077777778	66.4002777778	3
4	31.0744444444	53.3191666667	1

Cool but poor



This chatbot only sees the tabular data but is clueless about the metadata

# Tell me the range of latitudes and longitudes with the highest number of events

 HARVARD  
Dataverse

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

Harvard Dataverse > Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends >

## afginc.tab

This file is part of "Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends".

Version 1.0

File Citation

Koren, Ore, 2017, "afginc.tab", *Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends*, <https://doi.org/10.7910/DVN/PHHZI7/GVJOH3>, Harvard Dataverse, V1, UNF:6:nUddO4V8XEZZCnhI+9rxYg== [fileUNF]

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

Dataset Citation

Koren, Ore, 2017, "Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends", <https://doi.org/10.7910/DVN/PHHZI7>, Harvard Dataverse, V1, UNF:6:NDQ2oRK6QGGrA+OZgUMdDA== [fileUNF]

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

File Tools Metadata Versions

File Tools ▾ Open in New Window

Tell me what you want to know

longitudes with the highest number of events

Answer please

The range of latitudes and longitudes with the highest number of events is 30.9666666666667, 61.8833333333333 to 32.0, 67.25.


Unnamed: 0	latitude	longitude	event
1	30.9666666667	61.8833333333	2
2	31	64	3
3	31.0077777778	66.4002777778	3
4	31.0744444444	53.3191666667	1

ok-ish



This chatbot only sees the tabular data but is clueless about the metadata

# Map the range of latitudes and longitudes with the highest number of events to the names of countries

 HARVARD  
Dataverse

[Add Data](#) [Search](#) [About](#) [User Guide](#) [Support](#) [Sign Up](#) [Log In](#)

[Harvard Dataverse > Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends >](#)

## afginc.tab

This file is part of "Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends".

Version 1.0

File Citation

Koren, Ore, 2017, "afginc.tab", *Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends*, <https://doi.org/10.7910/DVN/PHHZI7/GVJOH3>, Harvard Dataverse, V1, UNF:6:nUddO4V8XEZZCnhi+9rxYg== [fileUNF]

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

Dataset Citation

Koren, Ore, 2017, "Why insurgents kill civilians in capital cities: A disaggregated analysis of mechanisms and trends", <https://doi.org/10.7910/DVN/PHHZI7>, Harvard Dataverse, V1, UNF:6:NDQ2oRK6QGGrA+OZgUMdDA== [fileUNF]

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

File Tools

[Metadata](#) [Versions](#)

File Tools

[Open in New Window](#)

Tell me what you want to know

map the range or latitudes and longitudes with the highest number of events to the names of countries

Answer please

The countries with the highest number of events in the given range of latitudes and longitudes are Afghanistan, Iran, and Pakistan.

Unnamed: 0	latitude	longitude	event
1	30.9666666667	61.8833333333	2
2	31	64	3
3	31.0077777778	66.4002777778	3
4	31.0744444444	53.3191666667	1

LLM kicks in



This chatbot only sees the tabular data but is clueless about the metadata

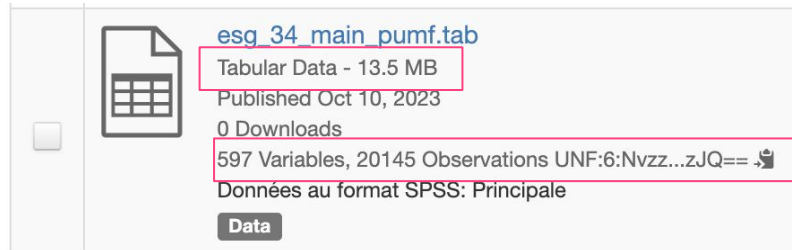
# Tabular Data Ingest

Tabular data formats (CSV, Excel, SPSS, Stata, R)

- Tabular data ingest
  - Conversion to Tab-delimited text
  - Metadata extraction of variables
  - Data Documentation Initiative (DDI) metadata standard

Reuses and integrations

- Enables variable metadata searching
- Preview and exploration
- Summary statistics and analysis
- Download as original format, Tab-delimited or R format



# Tabular data: depositing files

- Upload tabular files (SPSS, STATA, R, CSV, Excel)
- Data ingest tips:
  - Responses/ observations as rows
  - Variables/concepts as columns
  - No notes / improper formatting
  - Does not support multiple spreadsheets in one file
  - Excel (XLSX) is supported
  - CSV requires documentation
  - Encode in original language
  - Ingest errors can occur

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16
1	1.040e+14	4	26	104012	99	1	1973	2	1642	1040	1040	1	6	1	5	5
2	1.040e+14	1	19	104008	99	2	1943	1	99	99	99	1	1	6	4	5
3	1.040e+14	4	98	104003	99	1	1990	1	99	99	99	1	6	3	99	11
4	1.040e+14	2	98	104010	99	1	1983	2	1756	1040	1040	1	6	3	99	11
5	1.040e+14	1	18	104007	99	2	1927	1	99	99	99	1	1	6	4	1
6	1.040e+14	2	19	104007	99	1	1983	1	99	99	99	1	6	2	5	3
7	1.040e+14	4	15	104005	99	2	1970	1	99	99	99	1	1	2	4	6
8	1.040e+14	4	19	104006	99	1	1942	1	99	99	99	2	1	6	5	88
9	1.040e+14	4	15	104005	99	1	1965	2	1040	1276	1040	7	4	1	5	5
10	1.040e+14	7	15	104004	99	2	1955	1	99	99	99	1	1	2	5	5
11	1.040e+14	7	15	104003	99	2	7777	1	99	99	99	1	1	6	77	77
12	1.040e+14	4	18	104005	99	2	1938	1	99	99	99	1	3	6	5	5
13	1.040e+14	7	17	104005	99	1	1945	1	99	99	99	1	1	6	5	4
14	1.040e+14	4	18	104005	99	2	1949	2	1040	1380	1040	1	4	6	5	5
15	1.040e+14	2	15	104003	99	2	7777	1	99	99	99	1	1	2	4	5
16	1.040e+14	4	32	104007	99	1	1974	1	99	99	99	1	6	2	5	5
17	1.040e+14	4	98	104006	99	1	7777	1	99	99	99	1	6	3	99	11
18	1.040e+14	4	25	104010	99	2	1968	1	99	99	99	1	1	2	5	1
19	1.040e+14	4	40	104007	99	1	1967	1	99	99	99	1	1	2	6	2
20	1.040e+14	1	23	104004	99	1	1932	1	99	99	99	1	1	6	4	3
21	1.040e+14	2	18	104003	99	2	1965	1	99	99	99	1	6	2	5	5
22	1.040e+14	1	27	104011	99	1	1956	1	99	99	99	2	1	1	5	2
23	1.040e+14	4	18	104004	99	2	1923	1	99	99	99	1	3	6	3	3
24	1.040e+14	1	19	104006	99	2	1952	1	99	99	99	1	4	2	5	3
25	1.040e+14	3	16	104004	99	2	1947	2	1276	1040	1040	1	1	6	3	3

# Tabular data: variable metadata

- Assigns Unique identifiers
- Structured metadata for names, labels, codes, values
- Data contents:
  - Variables
  - Questionnaires & interviews (question text, universes)
  - Recodes, codes, values and notes
- Weighting
- Summary statistics
- Export variable metadata
- Export HTML and XML Codebook

Variable Description	
Variable Groups	
	<ul style="list-style-type: none"><li>• <a href="#">CAN: Cannabis</a></li><li>• <a href="#">DV: Derived Variables</a></li><li>• <a href="#">ALC: Alcohol</a></li><li>• <a href="#">IU: Initial use</a></li><li>• <a href="#">VAP: Vaping</a></li><li>• <a href="#">OTP: Other tobacco product status</a></li><li>• <a href="#">RD: Reference date</a></li><li>• <a href="#">DEM2: Demographics 2</a></li><li>• <a href="#">SRV: Survey Related Variables</a></li><li>• <a href="#">GDR: Gender</a></li><li>• <a href="#">TRC: Tobacco</a></li></ul>
CAN: Cannabis	
Variables within CAN: Cannabis	
	<ul style="list-style-type: none"><li>• <a href="#">Smoked cannabis - lifetime</a></li><li>• <a href="#">Age first time smoked cannabis</a></li><li>• <a href="#">Frequency smoked cannabis - past 30 days</a></li><li>• <a href="#">Nb of days smoked cannabis at least once past month - grouped</a></li><li>• <a href="#">Frequency smoked cannabis/tobacco mix - past 30 days</a></li><li>• <a href="#">Nb of days smoked can /toba mix at least once a week - past 30 days</a></li><li>• <a href="#">Frequency consumed edibles - past 30 days</a></li><li>• <a href="#">Vaped cannabis - lifetime</a></li><li>• <a href="#">Age first time vaped cannabis - grouped</a></li><li>• <a href="#">Frequency vaped cannabis - past 30 days</a></li><li>• <a href="#">Nb of days vaped cannabis at least once past month - 30 days - grouped</a></li><li>• <a href="#">Where get devices/liquids - Make your own vaping liquid</a></li><li>• <a href="#">Where get devices/liquids - Compassion club/dispensary/storefront</a></li><li>• <a href="#">Where get devices/liquids - Online</a></li><li>• <a href="#">Where get devices/liquids - Shared around a group of friends</a></li><li>• <a href="#">Where get devices/liquids - An acquaintance</a></li><li>• <a href="#">Where get devices/liquids - A family member</a></li><li>• <a href="#">Where get devices/liquids - A friend</a></li><li>• <a href="#">Where get devices/liquids - A dealer</a></li><li>• <a href="#">Where get devices/liquids - Other</a></li></ul>
DV: Derived Variables	
Variables within DV: Derived Variables	
	<ul style="list-style-type: none"><li>• <a href="#">Age group of person</a></li><li>• <a href="#">Province of residence (collection)</a></li><li>• <a href="#">Smoking Status</a></li><li>• <a href="#">Cannabis smoked in the last 30 days</a></li><li>• <a href="#">Vaped in the last 30 days</a></li><li>• <a href="#">Vaped cannabis in the last 30 days</a></li><li>• <a href="#">Consumed alcohol in the last 30 days</a></li><li>• <a href="#">First product tried</a></li></ul>

Source: Canadian Tobacco and Nicotine Survey, 2021.  
DDI Codebook (Odesi)

# Geospatial Data



[cd\\_2006\\_orig\\_CBF.zip](#)

Shapefile as ZIP Archive - 27.2 MB

Published Apr 4, 2023

0 Downloads

MD5: c0e...22e 

UNI-CEN Digital Boundary File for Census Division (CD), 2006, in Esri Shapefile format and NAD83 CSRS / EPSG:3348 projection. This file is part of the CBF-Original Shorelines series. Retrieved from: <https://borealisdata.ca/dataverse/unicen>. For more information about the project, visit <https://observatory.uwo.ca/unicen>.



[pr\\_1871\\_orig\\_CBF.geojson](#)

GeoJSON - 21.5 MB

Published Apr 4, 2023

2 Downloads

MD5: 14f...fa6 

UNI-CEN Digital Boundary File for Province/Territory (PR), 1871, in geojson format and WGS84 / EPSG:4326 projection. This file is part of the CBF-Original Shorelines series. Retrieved from: <https://borealisdata.ca/dataverse/unicen>. For more information about the project, visit <https://observatory.uwo.ca/unicen>.

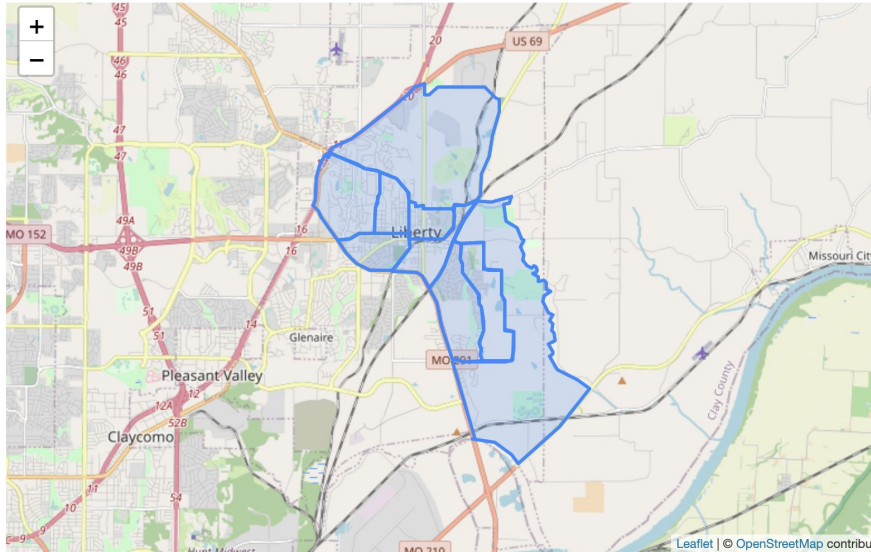




# GeoJSON Previewer

Preview Metadata Versions

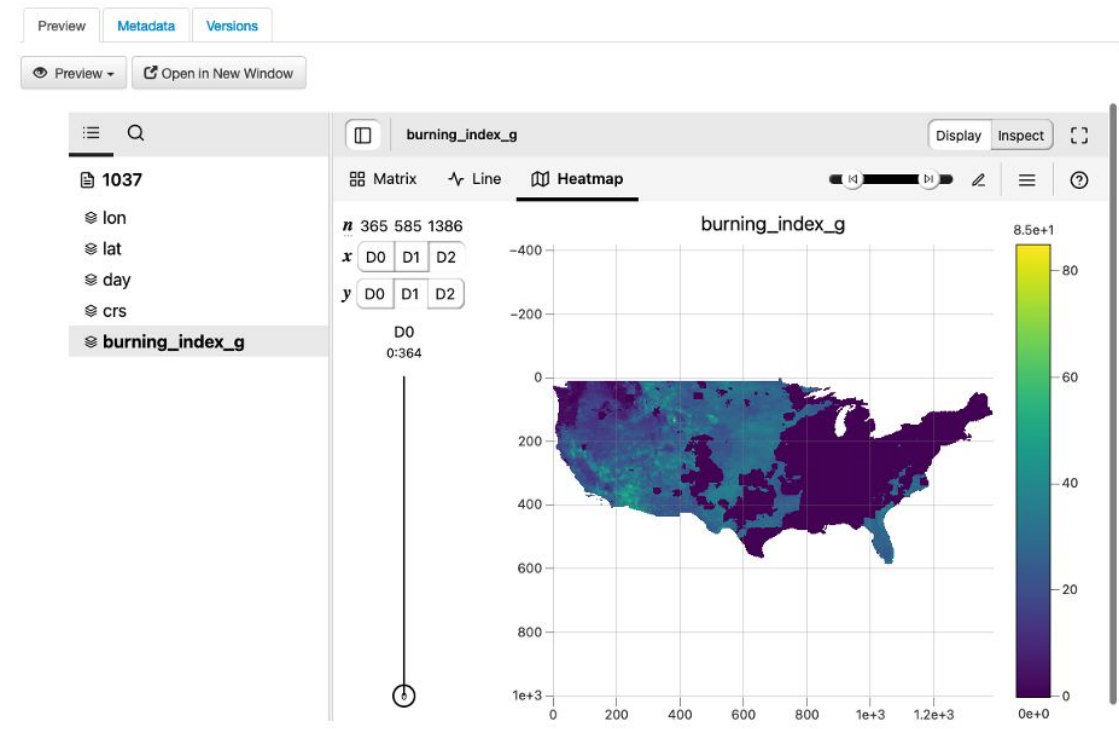
Open in New Window



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

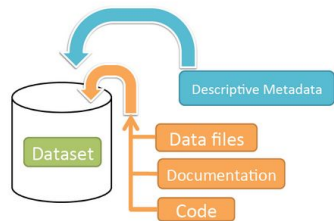
# NetCDF & HDF5 Previewer

- Common scientific data formats
- Metadata extracted in NcML (XML) format and saved as an auxiliary file
- H5Web visualization and file preview tool
  - Preview data
  - Explore and visualize data



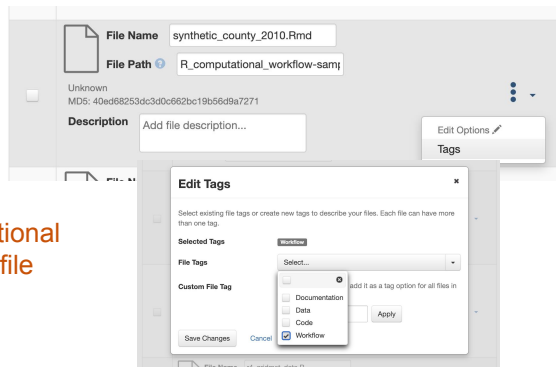
# File Types, Format, Documentation

Schematic Diagram of a Dataset in Dataverse 4.0

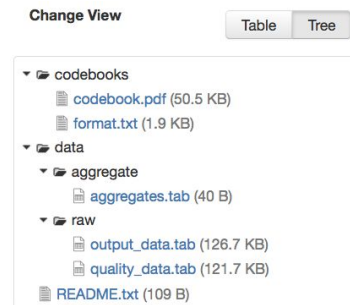


Container for your data, documentation, and code.

Computational  
workflow file  
support



Folder hierarchy support



Tabular data support

## Supported File Formats

Tabular Data ingest supports the following file formats:

File format	Versions supported
SPSS (PDR 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)

File level access control

## Access

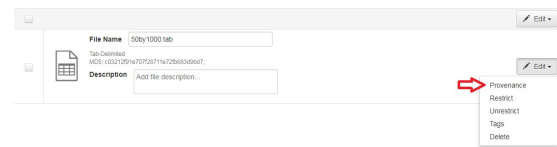
Public (1,784,339)

Restricted (50,591)

Embargoed then Public (154)

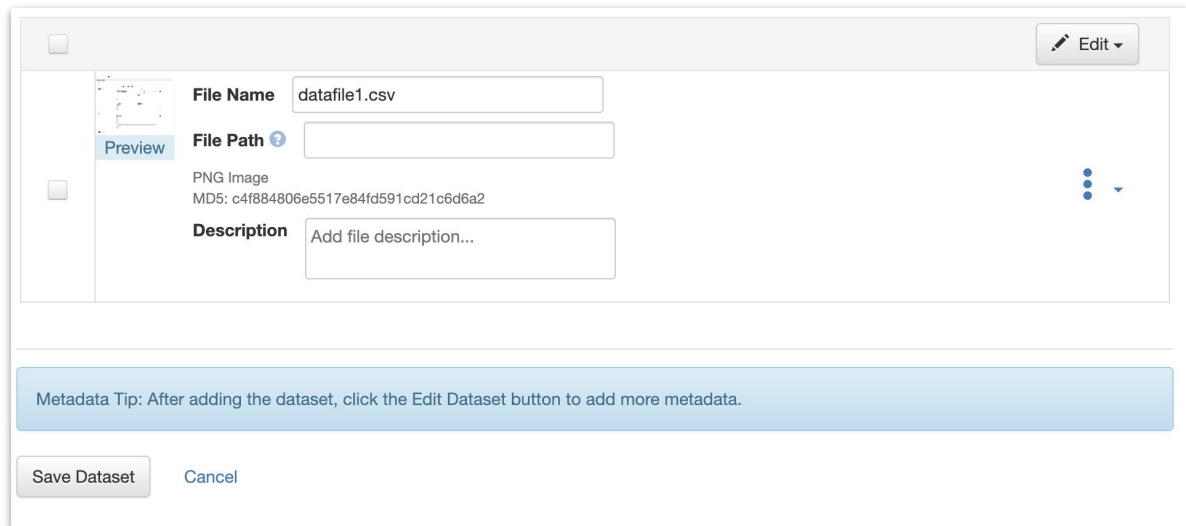
Embargoed then Restricted (14)

File level Provenance support



# Metadata you can use to describe files

- File names
- File paths
- Descriptions



The screenshot shows a web interface for editing file metadata. At the top right is an 'Edit' button with a pencil icon. The main area contains a 'Preview' tab on the left and a form on the right. The form has three fields: 'File Name' with the value 'datafile1.csv', 'File Path' with a question mark icon and an empty field, and 'Description' with the placeholder text 'Add file description...'. Below the 'File Path' field, it says 'PNG Image' and 'MD5: c4f884806e5517e84fd591cd21c6d6a2'. At the bottom, there is a blue tip box that says 'Metadata Tip: After adding the dataset, click the Edit Dataset button to add more metadata.' and two buttons: 'Save Dataset' and 'Cancel'.

File Name: datafile1.csv

File Path ?

Preview

PNG Image

MD5: c4f884806e5517e84fd591cd21c6d6a2

Description: Add file description...

Metadata Tip: After adding the dataset, click the Edit Dataset button to add more metadata.

Save Dataset Cancel


# Metadata you can use to describe files


- Tags
- Restrictions
- Embargo dates and reasons

### Restrict Access

Restricting limits access to published files. People who want to use the restricted files can request access by default. **If you disable request access, you must add information about access to the Terms of Access field.**

Learn about restricting files and dataset access in the [User Guide](#).

**Request Access**  ☒ Enable access request

**Terms of Access for Restricted Files** 

Save Changes [Cancel](#)

# Metadata you can use to describe files

- Tags
- Restrictions
- Embargo dates and reasons

## Edit Embargo

Edit the planned embargo for the selected file or files. Once this dataset version is published, you will need to contact an administrator to change the embargo date or reason of the file or files.

**Add or Change**      **Select the embargo end-date \***

2023-11-14

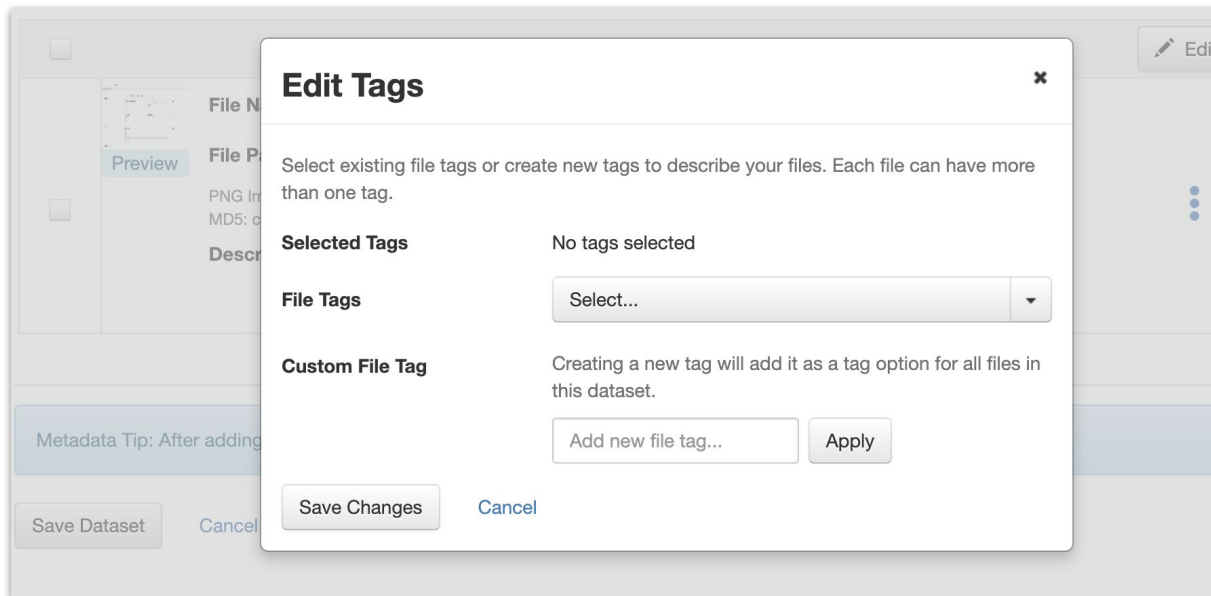
**November 2023**

S	M	T	W	T	F	S
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18

Save Changes

# Metadata you can use to describe files

- Tags
- Restrictions
- Embargo dates and reasons



# Metadata Dataverse creates for ingested tabular files

- Counts of variables and observations
- Variable statistics
- Checksums (UNFs)

File Metadata ^	
File UNF	UNF:6:HSeMI8h1XrpAPAOzE50Qow==
Original File MD5	05ed7a1c7330df3e88e57629bf15b5ff
Type	Tab-Delimited
Variables	8
Observations	1121

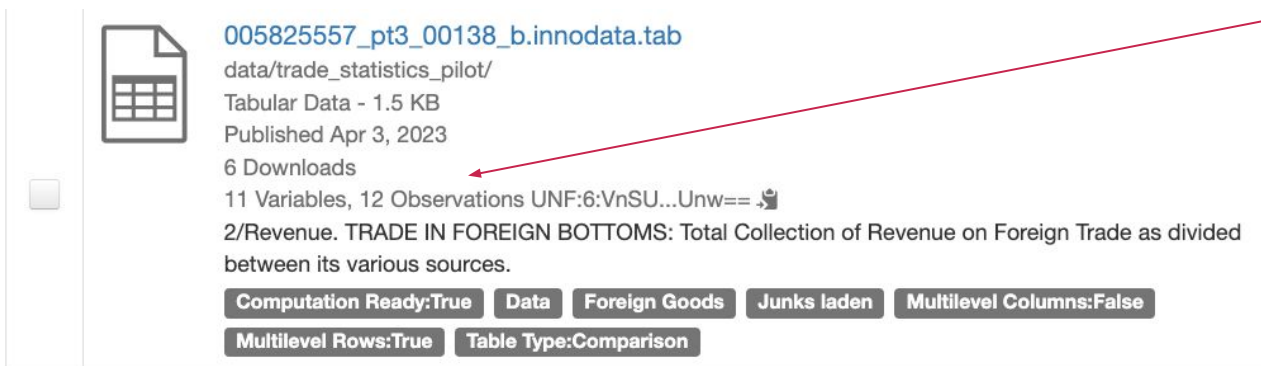
## TIME

f7570319 Location: Summary Statistics: Min. 2000.0; Valid 1121.0; Mean 2010.6030330062445; Max. 2021.0; StDev 6.318365357642705;

Variable Format: numeric



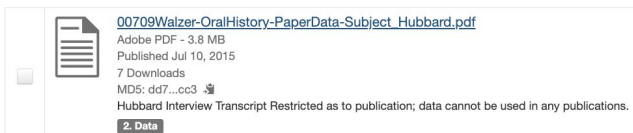
# File level metadata considerations and features



005825557\_pt3\_00138\_b.innodata.tab  
data/trade\_statistics\_pilot/  
Tabular Data - 1.5 KB  
Published Apr 3, 2023  
6 Downloads  
11 Variables, 12 Observations UNF:6:VnSU...Unw== .  
2/Revenue. TRADE IN FOREIGN BOTTOMS: Total Collection of Revenue on Foreign Trade as divided between its various sources.

Computation Ready:True Data Foreign Goods Junks laden Multilevel Columns:False  
Multilevel Rows:True Table Type:Comparison

Metadata extracted from tabular and other files with tabular ingest feature: Observations, Variables



00709Walzer-OralHistory-PaperData-Subject\_Hubbard.pdf  
Adobe PDF - 3.8 MB  
Published Jul 10, 2015  
7 Downloads  
MD5: dd7...cc3  
Hubbard Interview Transcript Restricted as to publication; data cannot be used in any publications.

2. Data

File level “terms”



005825557\_pt3\_00140.innodata.txt  
data/trade\_statistics\_pilot/  
Plain Text - 194 B  
Published Apr 3, 2023  
0 Downloads  
MD5: e18...5a8  
File associated with data tables series: Population:15

Customs Data JAMES H. HART STATISTICS

File level “tags”

Plan to populate additional metadata for files: description, edit file names, add tags

# The User Experience: Files and External Tools

# External Tools

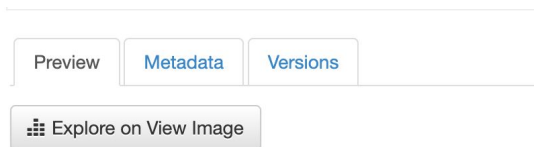
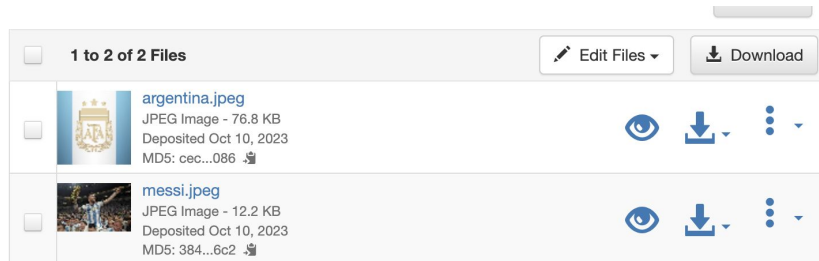
- External tools can provide additional features that are not part of the Dataverse Software itself, such as data file previews, visualization, and curation
- Communicate with Dataverse via its robust APIs
- Can be defined for Datasets or Files
- Can be Read-only or Read / Write

## Contents:

- [Inventory of External Tools](#)
- [Managing External Tools](#)
  - [Adding External Tools to a Dataverse Installation](#)
  - [Listing All External Tools in a Dataverse Installation](#)
  - [Showing an External Tool in a Dataverse Installation](#)
  - [Removing an External Tool From a Dataverse Installation](#)
- [Testing External Tools](#)
  - [File Level vs. Dataset Level](#)
  - [File Level Explore Tools](#)
  - [File Level Preview Tools](#)
  - [File Level Query Tools](#)
  - [File Level Configure Tools](#)
  - [Dataset Level Explore Tools](#)
  - [Dataset Level Configure Tools](#)
- [Writing Your Own External Tool](#)

# 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



# File Previewers (more examples)

Preview

Metadata

Versions

Explore on Read Document

Previous

Next

Page: 13 / 59

GEOMETRICAL SOLUTIONS DERIVED FROM MECHANICS.

ARCHIMEDES TO ERATOSTHENES, GREETING:

Some time ago I sent you some theorems I had discovered, writing down only the propositions because I wished you to find their demonstrations which had not been given. The propositions of the theorems which I sent you were the following:

1. If in a perpendicular prism with a parallelogram<sup>1</sup> for base a cylinder is inscribed which has its bases in the opposite parallelograms<sup>1</sup> and its surface touching the other planes of the prism, and if a plane is passed through the center of the circle that is the base of the cylinder and one side of the square lying in the opposite plane, then that plane will cut off from the cylinder a section which is bounded by two planes, the intersecting plane and the one in which the base of the cylinder lies, and also by as much of the surface of the cylinder as lies between these same planes: and the detached

File Tools

Metadata

Versions

File Tools Open in New Window

	NameFootballPlayer	International
1	Pedro Bonifacio Suárez Pérez \"Arico\" (Pedro Suárez)	Argentina
2	Milovan \"El Grande Milovan\" Jakšić	Yugoslavia
3	Ernest Libérati	France
4	Alexandre Villaplane (Captain)	France
5	Roberto Gayón Márquez	Mexico
6	Andrew \"Andy\" Auld	United States
7	James \"Jim\" Brown	United States
8	Jimmy Gallagher	United States
9	Alfred Eisenbeisser	Romania
10	Ladislau Raffinsky	Romania
11	Bartholomew \"Bertie\" or \"Bart\" McGhee (Bart McGhee)	United States
12	George Moorhouse	United States
13	Alexander \"Alec\" Wood	United States
14	Lorenzo Fernández	Uruguay
15	Constantino Urbieto Sosa	Argentina
16	Štefan Čambal	Czechoslovakia
17	Ferdinand Daučík	Czechoslovakia
18	Géza Kalocsay	Czechoslovakia
19	František Svoboda	Czechoslovakia
20	Matthias Sindelar (born as Matěj Šindelář)	Austria
21	Joseph Alcazar	France
22	Roger Courtois	France

# Zip File Previewer +

- A previewer that will show you the internal content and structure of a zip file (or electronic lab notebook)
- Uses the Range functionality in our Access api; so it's not just a viewer, it's an individual file unpacker

Preview

Metadata

Versions

Open in New Window

To download the complete zip file, please use the **Access File** button above.

▼	📁 BalsacGallery		
▼	📁 BalG_Adsorbates		
📄	Ag110EtO.eps	44.26 kB	↓
📄	Ag110EtO.jpg	147.77 kB	↓
📄	C2H4-C2H22.gif	98.23 kB	↓
📄	COO-CO2.gif	49.64 kB	↓
📄	Cu100KMr.eps	41.86 kB	↓
📄	Cu100KMr.jpg	145.94 kB	↓
📄	Cu100LiMr.eps	43.06 kB	↓
📄	Cu100LiMr.jpg	142.56 kB	↓
📄	Cu100NaMr.eps	41.86 kB	↓
📄	Cu100NaMr.jpg	145.92 kB	↓
📄	Cu110NHxO.eps	24.66 kB	↓
📄	Cu110NHxO.jpg	127.20 kB	↓
📄	Cu111+(4x4)-C60.eps	373.27 kB	↓
📄	Cu111+(4x4)-C60.jpg	379.56 kB	↓
📄	Cu111-C9H10allyl.eps	54.77 kB	↓
📄	Cu111-C9H10allyl.jpg	154.49 kB	↓
📄	Cu111-C9H10alpha.eps	54.78 kB	↓
📄	Cu111-C9H10alpha.jpg	159.73 kB	↓
📄	Cu111-C9H10trans.eps	54.78 kB	↓
📄	Cu111-C9H10trans.jpg	160.98 kB	↓

# File Exploration, Curation, and Query Tools

- File level **explore** tools provide a variety of features from data visualization to statistical analysis
- File level **configure** tools allow (authorized) users to send metadata about the file back to Dataverse
- 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

Forum Covid-19 Tracking

English

Forum\_COVID Tracking\_Data.tab

Forum Research Inc. 2020, "Forum Covid-19 Tracking", <https://doi.org/10.5683/SP2/YM8BCJ>, Borealis, V1, UNF:6:b2sqE84ecQkI2Y2CQgXUkA== [fileUNF]

< Hide Groups

Download

Save to Dataverse



















Add Group +

Search

Items per page 25

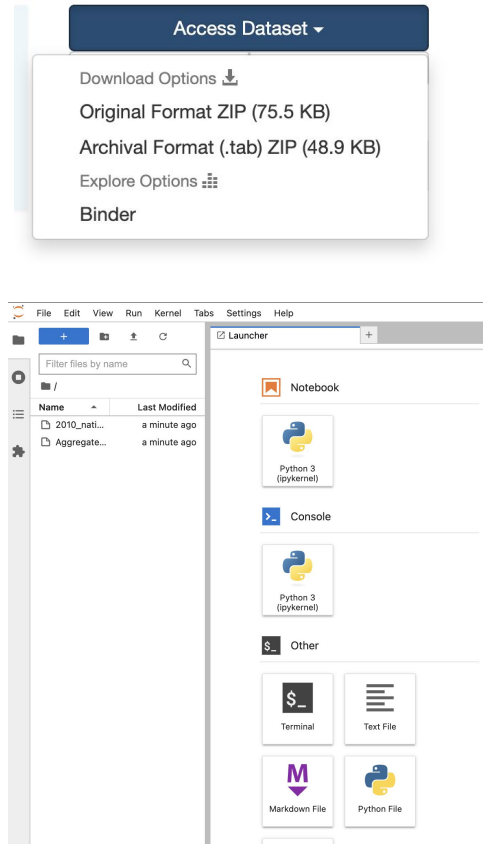
1 - 21 of 21

All 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?				 

# 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** (*coming soon*) - provides recommendations to improve the dataset metadata



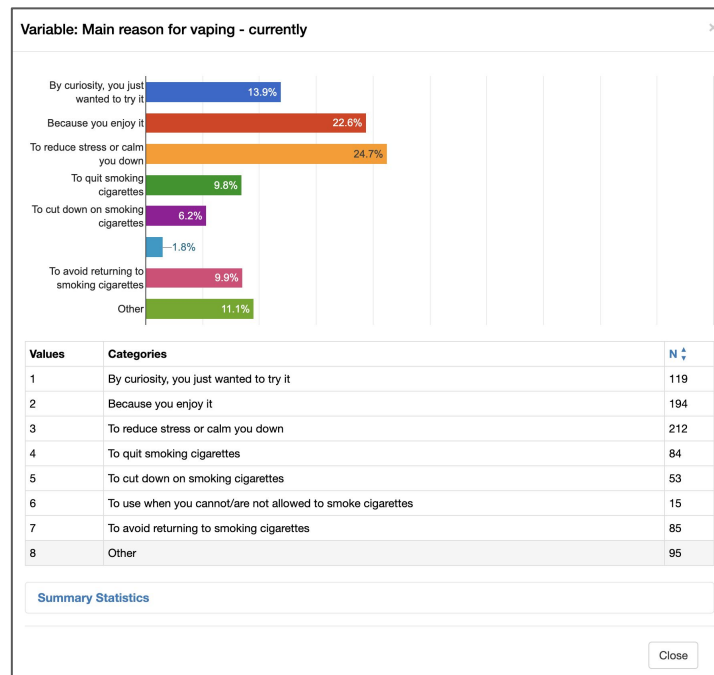


# Spotlight: Survey Data Analysis using Data Explorer

- Data Explorer 2.0 [Github](#) (open-source external Dataverse tool)
- Contributed by Borealis (Canada)

# Data Analysis using Data Explorer

- Helps users understand the data
  - Preview data and metadata for variables, questions, notes, universes
  - Get started with data analysis; digging, filtering, summary statistics
- Analyse and explore data
  - Frequencies charts
  - Cross-tabulation builder
  - Subsetting & download
- Analysing in Data Explorer promotes:
  - Data reuse
  - Transparency and reproducibility
  - Teaching data concepts
  - Open access and sharing



Source: VAP\_35R, Canadian Tobacco and Nicotine Survey, 2021.  
Odesi <https://odesi.ca>

# Connecting to Data Explorer



# Data Explorer

Search for  
variables

Cross tabulate  
two or more  
variables

View multiple  
summary  
statistics

Subset and  
download

Employment Insurance Coverage Survey, 2021 [Canada]

EICS-4428-E-2021\_F1.tab

Labour Statistics Division, 2023, "Employment Insurance Coverage Survey, 2021 [Canada]" <https://doi.org/10.5072/FK2/TTRIBT>, Borealis, V1, UNF:6:QeZj6HFqsjnX706WrMLdQ== [fileUNF]

< Hide Groups

All Variables

Search

Items per page: 25 1 - 25 of 65 < >

Confirm Type	ID	Name	Label	Weight	View Summary Statistics	View Categories	View Questions
Changes in Income	<input type="checkbox"/>	v6012811	PUMFID	Sequential Identification Number	v6012809		
Parents	<input type="checkbox"/>	v6012804	SRYP	Survey reference year	v6012809		
Benefits	<input type="checkbox"/>	v6012769	REGION6	Province or region			
Unemployment	<input type="checkbox"/>	v6012819	SEX	Sex of respondent			
Weight Variable	<input type="checkbox"/>	v6012801	AGECAT	Age of respondent - grouped			
Sociodemographics	<input type="checkbox"/>	v6012797	EDUC	Highest educational attainment			
Additional Sources of Funding	<input type="checkbox"/>	v6012779	TYPES	Survey type as determined for the EICS	v6012809		
Job Information	<input type="checkbox"/>	v6012822	WORKNOW	Respondent currently working at a job or business	v6012809		
Parental benefits for spouses	<input type="checkbox"/>						

Browse by  
Variable  
Groups

Select /  
Unselect  
Variables

View summary statistics  
, frequencies, question  
information

# Summary Statistics

Employment Insurance Coverage Survey, 2021 [Canada]

EICS-4428-E-2021\_F1.tab

Labour Statistics Division, 2023, "Employment Insurance Coverage Survey, 2021 [Canada]", <https://doi.org/10.5072/FK2/TTRiBT>, Borealis, V1, UNF-6:QUEzi6HFasinX706WMLdQ== [fileUNF]

< Hide Groups

All Variables Search

Confirm Type

Changes in Income

Parents

Benefits

Unemployment

Weight Variable

Sociodemographics

Additional Sources of Funding

Job Information

Parental benefits for spouses

Unassigned

Economic Family

WORKNOW: Respondent currently working at a job or business

Values	Categories	Count	Count Percentage(%)	Weighted Count	Weighted Count Percentage(%)
6	Valid skip	3,124	44.124	3,361,659.873	45.113
1	Yes	1,866	26.356	1,940,991.525	26.048
8	Refusal	0	0	0	0
2	No	2,074	29.294	2,129,783.673	28.582
7	Don't know	0	0	0	0
9	Not stated	16	0.226	19,147.644	0.257

Summary Statistics

Download

Items per page 25 1 - 25 of 65

Weight	View Summary Statistics	View Categories	View Questions
v6012809			
v6012809			
v6012809			
v6012809			
v6012809			

Identifies if a person is a landed immigrant

# Analyse data

## Select Variables For Cross Tabulation

VAP\_15AR - Nb of days vaped e-liquid with nicotine - pas ☐

VAP\_15BR - Nb of days vaped e-liquid without nicotine - ☐

VAP\_15CR - Nb of days vaped e-liquid didn't know what it ☐

VAP\_20R - Nb of times pick up vaping device - Day respo ☐

VAP\_21R - Nb of puffs taken each time vaping device is ☐

VAP\_30R - Flavour respondent vape most often ☐

VAP\_40AR - Where get vaping devices - Vape shop ☐

VAP\_40BR - Where get vaping devices - Convenience store/ ☐

VAP\_40CR - Where get vaping devices - Supermarket/grocer ☐

VAP\_40DR - Where get vaping devices - Online ☐



Rows

VAP\_35R - Main reason for vaping - currently ☐

Columns

AGEGROUP - Age group of person ☐



OK

## Cross Tabulation Table

AGEGROUP - Age group of person	15 to 19 years old	20 to 24 years old	25 to 34 years old	35 to 44 years old	45 to 54 years old	55 to 64 years old	65 years old and older	Valid skip	Don't know	Refusal	Not stated	Total
<b>VAP_35R - Main reason for vaping - currently</b>												
By curiosity, you just wanted to try it	59 0.6%	45 0.45%	7 0.07%	3 0.03%	3 0.03%	1 0.01%	1 0.01%	0 0%	0 0%	0 0%	0 0%	119 1.2%
Because you enjoy it	72 0.73%	102 1.03%	9 0.09%	7 0.07%	2 0.02%	1 0.01%	1 0.01%	0 0%	0 0%	0 0%	0 0%	194 1.96%
To reduce stress or calm you down	87 0.88%	105 1.06%	7 0.07%	8 0.08%	2 0.02%	0 0%	3 0.03%	0 0%	0 0%	0 0%	0 0%	212 2.14%
To quit smoking cigarettes	6 0.06%	37 0.37%	10 0.1%	11 0.11%	12 0.12%	2 0.02%	6 0.06%	0 0%	0 0%	0 0%	0 0%	84 0.85%
To cut down on smoking cigarettes	1 0.01%	28 0.28%	4 0.04%	8 0.08%	3 0.03%	5 0.05%	4 0.04%	0 0%	0 0%	0 0%	0 0%	53 0.53%
To use when you cannot/are not allowed to smoke cigarettes	2 0.02%	8 0.08%	0 0%	2 0.02%	2 0.02%	1 0.01%	0 0%	0 0%	0 0%	0 0%	0 0%	15 0.15%
To avoid returning to smoking cigarettes	2 0.02%	41 0.41%	15 0.15%	8 0.08%	12 0.12%	3 0.03%	4 0.04%	0 0%	0 0%	0 0%	0 0%	85 0.86%
Other	31 0.31%	49 0.49%	3 0.03%	6 0.06%	2 0.02%	2 0.02%	2 0.02%	0 0%	0 0%	0 0%	0 0%	95 0.96%
	1775	1087	641	612	841	1100	1764	0	0	0	0	6070

# Download and subset

Canadian Tobacco and Nicotine Survey, 2021

CTNS-5305-E-2021\_F1.tab

Statistics Canada, 2023, "Canadian Tobacco and Nicotine Survey, 2021", <https://doi.org/10.5072/FK2/ZHT5KX>, Borealis, V1, UNF:6:15CjFBAeehXclm89HK6dpA== [fileUNF]

< Hide Groups

Cross Tabulation Summary

Download

Download subset

Download original file

Download tab-delimited file

Download RData format file

Download variable metadata

All Variables

Search

Items per page 25

	ID	Name	Label	Weight
CAN: Cannabis				
DV: Derived Variables				
ALC: Alcohol				
IU: Initial use				
VAP: Vaping				
OTP: Other tobacco product status				
RD: Reference date				
DEM2: Demographics 2				
SRV: Survey Related Variables				
GDR: Gender				
TBC: Tobacco				
<input type="checkbox"/>	v6423547	AGEGROUP	Age group of person	
<input checked="" type="checkbox"/>	v6423552	ALC_05	Frequency drank alcoholic beverage - past 30 days	
<input checked="" type="checkbox"/>	v6423564	ALC_10	Freq drank 4 drinks or more on one occasion - past 12 months	
<input checked="" type="checkbox"/>	v6423571	CAN_05AR	Smoked cannabis - lifetime	
<input checked="" type="checkbox"/>	v6423570	CAN_05BR	Age first time smoked cannabis	
<input type="checkbox"/>	v6423525	CAN_10AR	Frequency smoked cannabis - past 30 days	
<input type="checkbox"/>	v6423539	CAN_10BR	Nb of days smoked cannabis at least once past month - grouped	
<input type="checkbox"/>	v6423524	CAN_15AR	Frequency smoked cannabis/tobacco mix - past 30 days	
<input type="checkbox"/>	v6423569	CAN_15BR	Nb of days smoked can./toba. mix at least once a week - past 30 days	
<input type="checkbox"/>	v6423529	CAN_17R	Frequency consumed edibles - past 30 days	

# Further Integrations

- Depositing and sharing tabular data in Dataverse
  - Improvements across tabular data ingest and Data Explorer (weighting cross-tabs, exporting charts, more visualizations, Data Documentation Initiative (DDI) support)
- Using Dataverse APIs you can build your own data exploration tools
- Data Explorer is an open, ready to use tool
- Data Explorer can be built into integrated data search sites (e.g.

## Data catalogs

The screenshot displays the Odesi data catalog interface. At the top, there's a header with the Odesi logo and language options (Français, English). Navigation links include 'My List (0)', 'Browse', 'Contact', and 'Deposit Data'. The search bar contains 'stress' and a dropdown menu for 'Variable'. Below the search bar, there are filters for 'Collection' (All collections selected), 'Sort' (Relevance), and 'Date (YYYY)' (From, To). A 'SEARCH' button is present. On the right, a text box describes Odesi as a Canadian social science data repository and online exploration and analysis tool, containing 5,700+ datasets curated by academic libraries in Canada. The main content area shows search results for 'stress', with 'Results 1 to 20 of 513' and 'Page 1 of 26'. The first result is 'Enquête sociale générale 2010, Cycle 24 [Canada]: Bien-être et stress lié au manque de temps, Fichier principal'. The second result is 'General Social Survey, Cycle 24, 2010 [Canada]: Time-Stress and Well-Being, Main File [version 5]'. On the right side of the results, there's a 'Limit your search:' section with a list of collections: Odesi 291, DU 235, Dlimt 184, POP 52, Borealis 38, and More. Below this, there's a 'Series' section with 'Canadian Tobacco Use Monitoring Survey 40', 'Enquête De Surveillance De L'usage Du Tabac Au Canada 34', and 'Canadian Community Health Survey 24'.



# The User Experience: Permissions and Access

# Permissions & Access: Flexible Permission System

- Supports multiple workflows by controlling who can deposit files on your behalf, and who can download files
- 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)

## Access control:

- Custom terms
- Restricted data must clearly define terms of access
- “Request access” workflow

The screenshot displays a web interface for managing permissions and access. It is divided into two main sections: 'Users/Groups' and 'Restricted Files'.

**Users/Groups Section:**

- Header: 'Users/Groups' with a dropdown arrow. Sub-header: 'All the users and groups that have access to restricted files in this dataset.'
- Buttons: 'Grant Access to Users/Groups' (with a user icon).
- Checkbox: 'Include Deleted Files' (checked).
- Count: '0 Users/Groups'.
- Table:

User/Group Name (Affiliation)	ID	Files	Access
There are no users or groups with access to the restricted files in this dataset.			










**Restricted Files Section:**

- Header: 'Restricted Files' with a dropdown arrow. Sub-header: 'All the restricted files in this dataset.'
- Checkbox: 'Include Deleted Files' (checked).
- Count: '0 Restricted Files'.
- Table:

File Name	Published	Users/Groups	Access
There are no restricted files in this dataset.			

**File List Section:**

- Filters: 'File Type: All' and 'Access: All' (both with dropdown arrows).
- Buttons: 'Download' and 'Request Access'.
- Count: '1 to 10 of 55 Files'.
- Table:

<input type="checkbox"/>	 <b>boxplot_2d.py</b> Python Source Code - 4.5 KB Published Jul 13, 2023 0 Downloads MD5: 093...1c9 	
<input type="checkbox"/>	 <b>converting_TSIS-1_to_UVspec_compatible_format.ipynb</b> Jupyter Notebook - 72.8 KB Published Jul 13, 2023 0 Downloads MD5: 6be...0d1 	
<input type="checkbox"/>	 <b>convolution-1.py</b> Python Source Code - 4.5 KB Published Jul 14, 2023 0 Downloads MD5: eb2...304 	

# Permissions & Access: Flexible Permission System

- Supports multiple workflows by controlling who can deposit files on your behalf, and who can download files
- 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)

## Access control:

- Custom terms
- Restricted data must clearly define terms of access
- “Request access” workflow
- Public, Restricted, Embargoed

## Access

Public (17,183)

Restricted (275)

Embargoed then Public (2)

Access: EmbargoedThenPublic ✕

1 to 2 of 2 Results

2-neu.txt

Draft

Unpublished

Embargoed



Mar 16, 2023 - [gezippt](#)

Plain Text - 9 B - MD5: ca8...198

text text

7610-how-to-use-multilingual-external-module (1).docx

Embargoed



Mar 16, 2023 - [Test dataset](#)

MS Word - 239.0 KB - MD5: 1bc...840

# Permissions & Access: Flexible Permission System

- Supports multiple workflows by controlling who can deposit files on your behalf, and who can download files
- 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)

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


CC0 1.0


- CC BY-NC-SA 4.0
- CC BY-ND 4.0
- CC BY-SA 4.0
- PDDL-1.0
- ODC-By 1.0
- ODbL 1.0
- OGL UK 3.0
- Custom Dataset Terms

### Restrict Access

Restricting limits access to published files. People who want to use the restricted files can request access by default. If you disable request access, you must add information about access to the Terms of Access field.

Learn about restricting files and dataset access in the [User Guide](#).



**Request Access**  ☒ Enable access request

**Terms of Access for Restricted Files** 

## Access control:

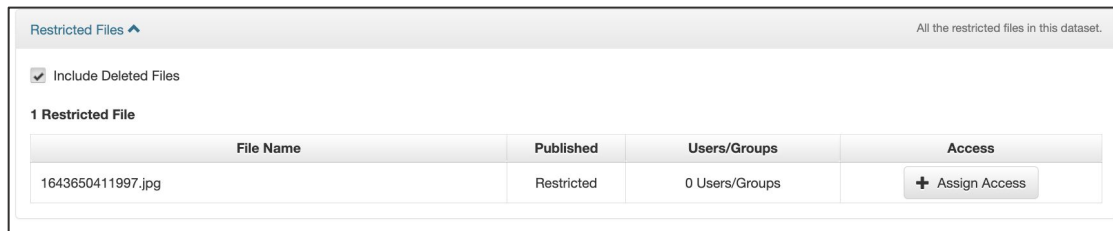
- Custom terms
- Restricted data must clearly define terms of access
- “Request access” workflow
- Public, Restricted, Embargoed
- Licenses, Custom Licenses

### Restricted Files + Terms of Access

<b>Restricted Files</b> 	There is 1 restricted file in this dataset.
<b>Request Access</b> 	Users may request access to files.

# Permissions & Access: Flexible Permission System

- Supports multiple workflows by controlling who can deposit files on your behalf, and who can download files
- 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)



Restricted Files ^

All the restricted files in this dataset.

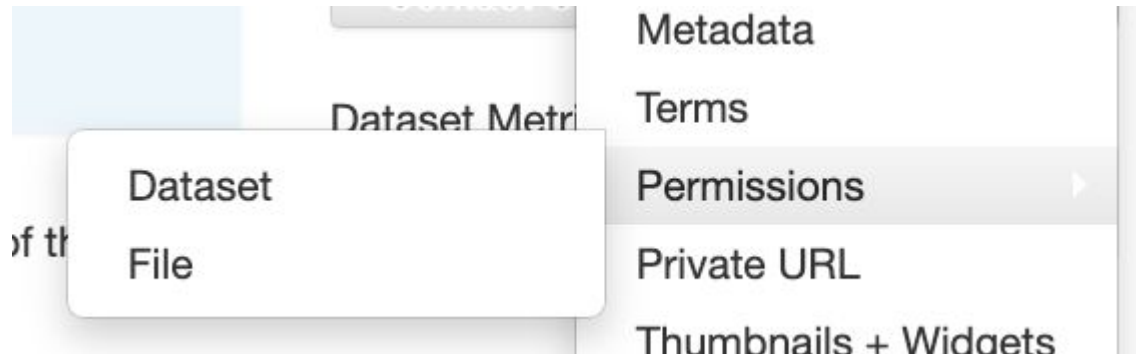
☒ Include Deleted Files

1 Restricted File

File Name	Published	Users/Groups	Access
1643650411997.jpg	Restricted	0 Users/Groups	<a href="#">+ Assign Access</a>

## Access control:

- Custom terms
- Restricted data must clearly define terms of access
- “Request access” workflow
- Public, Restricted, Embargoed
- Licenses, Custom Licenses
- Dataset or File level download options



# Permissions & Access: Flexible Permission System

- Supports multiple workflows by controlling who can deposit files on your behalf, and who can download files
- 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)

## Access control:

- Custom terms
- Restricted data must clearly define terms of access
- “Request access” workflow
- Public, Restricted, Embargoed
- Licenses, Custom Licenses
- Dataset or File level download options

## Prueba Demo Consorcio

**Draft** **Unpublished**

Saavedra, Paula, 2023, "Prueba De DRAFT VERSION" [?](#)

[Cite Dataset](#) [Learn about](#)



**Description** [?](#) pruebas 2

**Subject** [?](#) Agricultural Science

**License/Data Use Agreement**  [CC0](#)

**Publish Dataset** ▼

**Edit Dataset** ▼

**Contact Owner**

**Dataset Metrics**

**0 Downloads**

**Files (Upload)**

**Metadata**

**Terms**

**Permissions** ▶

**Private URL**

**Thumbnails + Widgets**

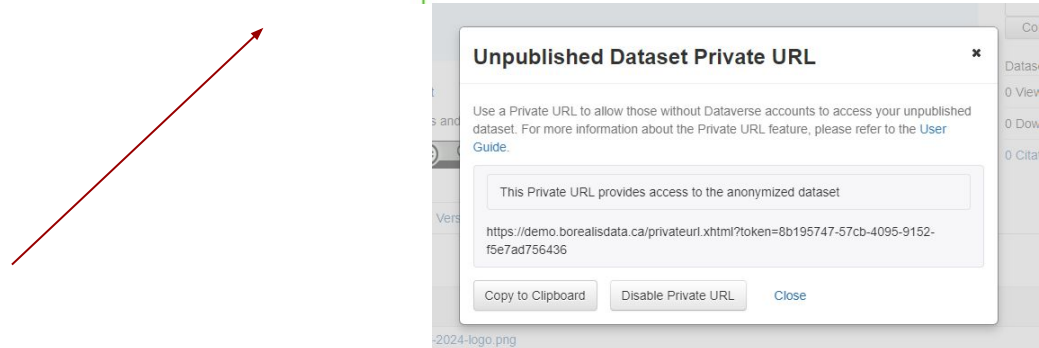
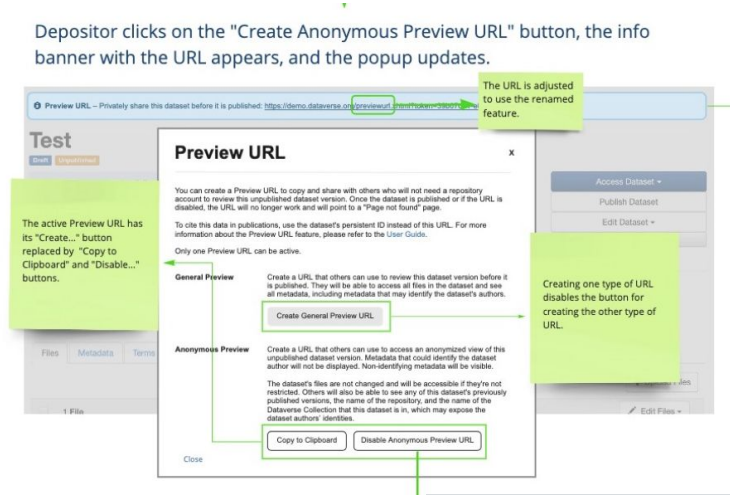
**Delete Dataset**

# Permissions & Access: Flexible Permission System

- Supports multiple workflows by controlling who can deposit files on your behalf, and who can download files
- 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)

## Access control:

- Custom terms
- Restricted data must clearly define terms of access
- "Request access" workflow
- Public, Restricted, Embargoed
- Licenses, Custom Licenses
- Dataset or File level download options



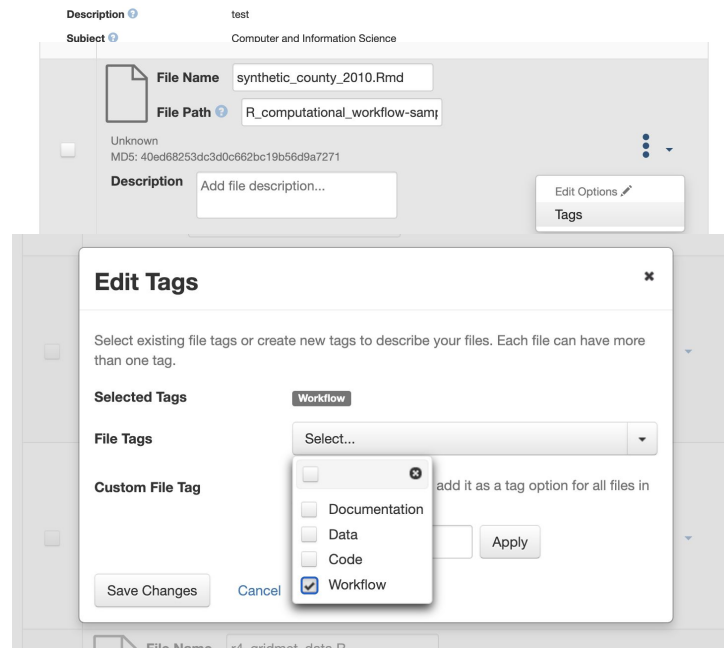
<https://github.com/IQSS/dataverse/issues/8185>

# The User Experience: Data Exploration



# Computational workflow and files support

- External tools exist to support **workflows** and **reproducibility**
- specific metadata to better support discoverability (e.g. the new “Dataset Feature” facet)
- Automatic checksum validation on **BagIt** file upload (based on the BagIt manifest)
- **Binder** support (to run Jupyter notebooks)  
[soon a connection to NERC/MOC]



# Dataverse & BagIt

- Generates BagIt zip file with complete metadata & all datafiles for a Dataverse dataset
- Conforms to RDA recommendations & includes complete JSON-LD/RDF metadata using [OAI-ORE](#) structure
- Imports BagIt packages as datasets, providing round-trip, export/import capability.

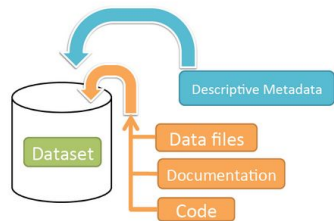
## Dataverse Archives Using BagIt

- [Texas Data Repository](#)  
(Duracloud/Chronopolis)
- [Qualitative Data Repository](#)  
(Google Cloud)
- [Harvard Data Commons](#)  
Workflow developed for depositing archival bags into the Harvard DRS repo (not in production yet)



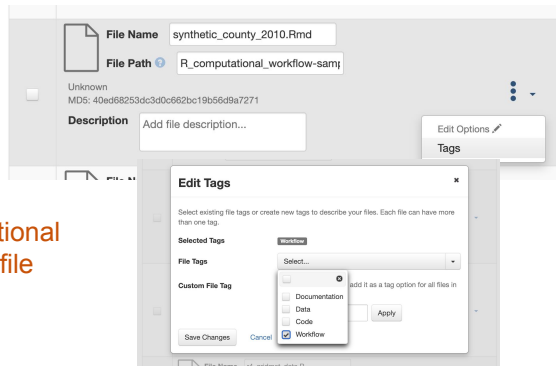
# File Types, Format, Documentation

Schematic Diagram of a Dataset in Dataverse 4.0

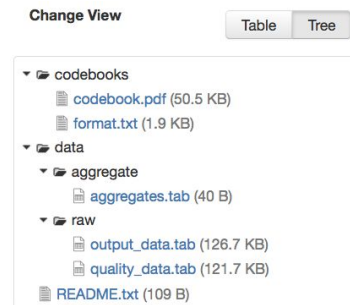


Container for your data, documentation, and code.

Computational workflow file support



Folder hierarchy support



Tabular data support

## Supported File Formats

Tabular Data ingest supports the following file formats:

File format	Versions supported
SPSS (PDR 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)

File level access control

## Access

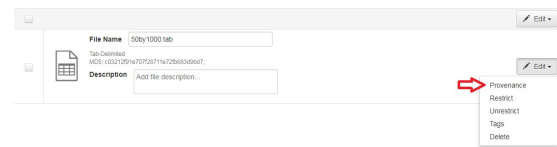
Public (1,784,339)

Restricted (50,591)

Embargoed then Public (154)

Embargoed then Restricted (14)

File level Provenance support



# The User Experience: 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)

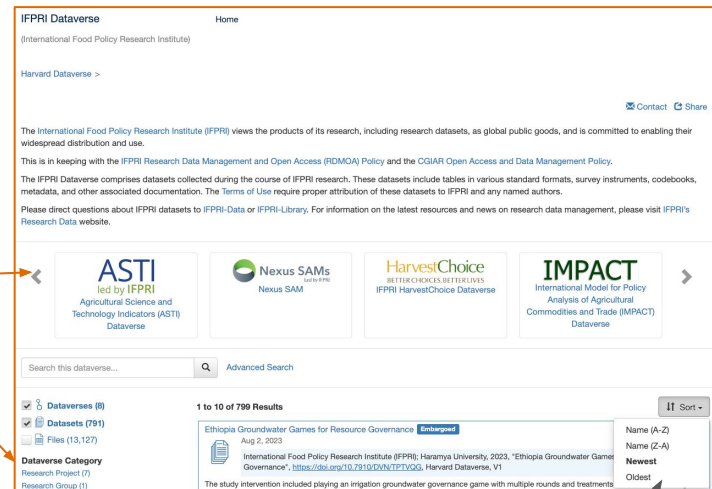
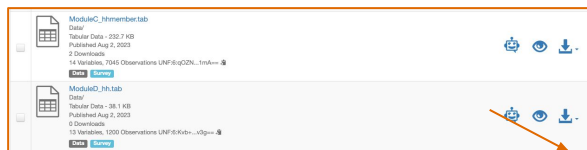


### Dataset details:

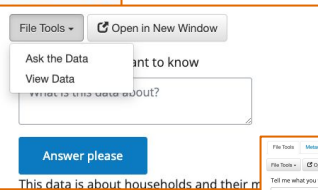
- Detailed dataset page
- Metadata
- Description
- Authors
- Citation
- Download options
- Documentation, readme, code...
- Multiple format download options

### Browse:

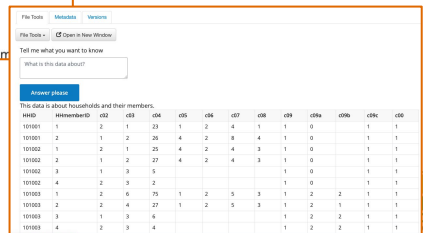
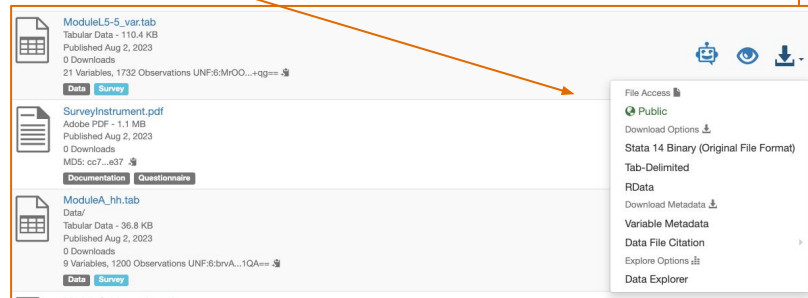
- Categories/Subjects
- Featured Datasets/collections



Ask the Data



Browse: Recently added



# 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>

## References:

Dataverse - [Goals, Roadmaps, and Releases](#)

Introduction to [Metadata in Dataverse Repositories](#): For  
Researchers and Support Staff

Introduction to [Data Handling on the Dataverse](#) Platform

Dataverse Software - [New Features and Future Plans](#)