

# **Data Handling on The Dataverse Project**

Sonia Barbosa, Gustavo Durand, Amber Leahey



Sonia Barbosa  
Manager, Harvard Dataverse  
Repository and Murray Archive

The Dataverse Project  
Data Science, IQSS  
Harvard University



Gustavo Durand  
Technical Lead and Architect  
of Dataverse

The Dataverse Project  
Data Science, IQSS  
Harvard University



Amber Leahey  
Service Director, Borealis, The  
Canadian Dataverse Repository /  
Le dépôt Dataverse canadien

Scholars Portal /University of  
Toronto Libraries

# Introduction

# Agenda

- Introduction to Dataverse
- Data Management Principles
- File Handling & Deposit
- Permissions and access
- Overview of File Handling Features
- Management of Various Data Types
- File Versioning
- Files and External Tools / Integrations
- Spotlight: Analysing Survey Data Using Data Explorer
- Thank you & Upcoming Events

# Introduction to Dataverse

## 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?](#)

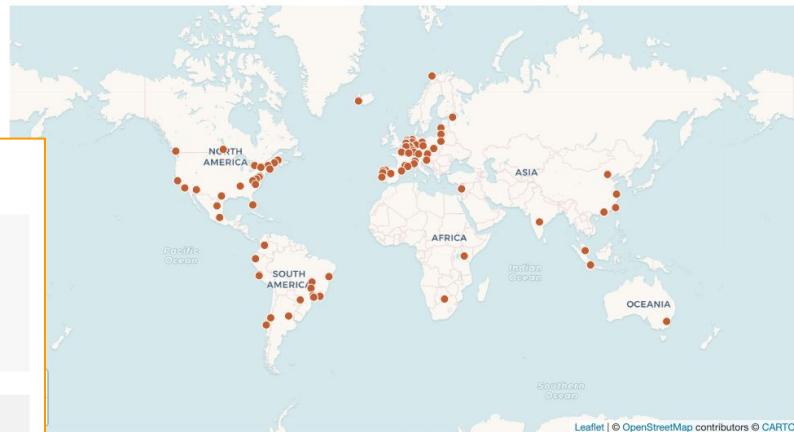


Developers

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?](#)

DATVERSE REPOSITORIES - A WORLD VIEW

105 Installations



# **Data Management Principles**

# Data Management Principles

- **Data Governance:** Data Ownership, Stewardship, Policies, Compliance, Ethics
- **Data Quality:** Accuracy, Consistency, Completeness, Reliability, Timeliness, Validity
- **Data Security:** Access control, Encryption, Authentication, Authorization, Auditing and Monitoring, Data Masking
- **Data Privacy:** Consent Management, PII protection, GDPR, HIPAA compliance
- **Data Lifecycle management:** Creation, Storage, Archiving, Disposal, Retention, Versioning
- **Data Integration:** ETL (Extract, Transform, Load), Data Connectors, Data Warehouse, Data Transformation, Data Pipelines
- **Data management:** Consolidation, synchronization, deduplication, governance
- **Metadata management:** Data description, data lineage, data catalog, data dictionary, data taxonomy, data versioning
- **Data backup and recovery:** Regular backup, disaster recovery plan, backup and recovery testing, business continuity plan, data redundancy
- **Data access and retrieval:** Querying/reporting, access APIs, Visualizations, self service and search capabilities
- **Data monitoring and auditing:** Real-time monitoring, Log analysis, Compliance checks, Anomaly detection, Audit trails, Security incident response
- **Data documentation:** Data models, Data schemas, Data dictionaries, Data lineage diagrams, Data change history, Data quality reports

# Data Management Principles and Dataverse

**Data Organization:** Repository access, folder support, and storage for different types of data files and associated metadata.

**Data Documentation:** Proper metadata and documentation

**Metadata Management:** Comprehensive metadata to describe their datasets (study level and file level).

**Data Sharing:** Publish datasets, assign persistent identifiers (DOIs), and specify access rights.

**Data Security:** Controlling access to data through permissions and embargos to ensure data security and compliance with privacy regulations.

**Data Preservation:** Long-term data preservation, ensuring that data remains accessible and usable over time.

While providing training and education on maintaining data quality, and providing tools for interoperability and integration.



# **File Handling and Deposits - Dataverse**

# Introduction to File Handling & Deposit - Dataverse

- Know **what to share**:
  - Research projects
  - Scientific data outputs
- **Organize files** for deposit
- File **types** and **formats**
- File **upload** options
- File **sizes**
- File-level **metadata support**
  - File naming, file descriptions, file-level tags, file-level edits

# File types and formats

## Text Data Formats:

- Plain Text
- CSV

## Tabular Data Formats:

- CSV (Comma-Separated Values)
- TSV (Tab-Separated Values)
- XLSX (Microsoft Excel)
- ODS (OpenDocument Spreadsheet)
- Excel CSV (Excel-Formatted CSV)
- SPSS
- Stata
- Parquet (for large files)

## Geospatial Data Formats:

- GeoJSON: A format for encoding geospatial data using JSON.
- Shapefile: A widely used format for geospatial vector data.

## Scientific Data Formats:

- NetCDF (Network Common Data Form): A format for storing scientific data, commonly used in environmental and climate sciences.
- HDF5 (Hierarchical Data Format): A versatile data format used in various scientific fields.

## Image Formats:

- PNG (Portable Network Graphics): A widely supported format for lossless image compression.
- SVG (Scalable Vector Graphics): An XML-based format for vector graphics.

## Audio and Video Formats:

- OGG (Ogg Vorbis): Open audio compression format.
- WebM: An open video format designed for the web.

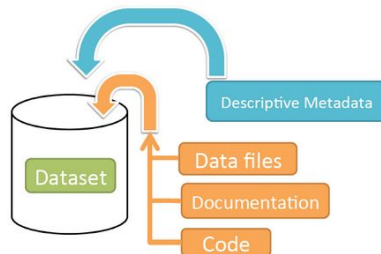
## Compression Formats:

- GZIP: Open compression format for files and data streams.
- 7-Zip: An open, high-compression archive format.

# Organize files for deposit

- Use descriptive file names
- Use a clear folder structure
- Use standard file formats
- Include documentation
- Use version control
- Clean the data
- Include metadata
- Use folder subdirectories
- Avoid special characters
- Use consistent file naming and formatting
- Partition if possible
- Ensure security and privacy
- Set appropriate file level permissions and access control
- Test what you have shared and ensure files open and download as expected
- Provide your contact information for your data users
- Specify the terms and conditions for reuse

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



Container for your data, documentation, and code.

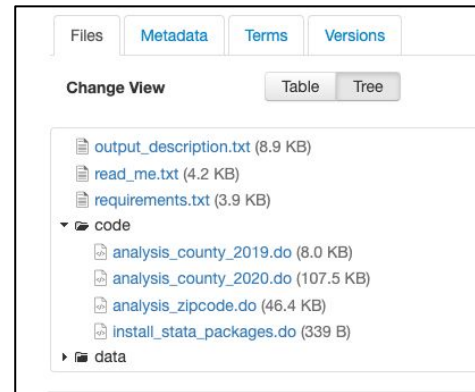
## Access

Public (1,784,339)

Restricted (50,591)

Embargoed then Public (154)

Embargoed then Restricted (14)



The screenshot shows the 'Files' tab of a dataset in Dataverse. The 'Change View' section has 'Table' selected. The file list includes:

File Name	Size	Download Icon
BalsacGallery		
Ag110EtO.eps	44.26 kB	Download
Ag110EtO.jpg	147.77 kB	Download
C2H4-C2H22.gif	98.23 kB	Download
COO-CO2.gif	49.64 kB	Download
Cu100KMr.eps	41.86 kB	Download
Cu100LMr.jpg	145.94 kB	Download
Cu100LMr.eps	43.06 kB	Download
Cu100LMr.jpg	142.56 kB	Download
Cu100NaMr.eps	41.86 kB	Download
Cu100NaMr.jpg	145.92 kB	Download
Cu110NtHtO.eps	24.66 kB	Download
Cu110NtHtO.jpg	127.20 kB	Download
Cu111+(4x4)-C60.eps	373.27 kB	Download
Cu111+(4x4)-C60.jpg	379.56 kB	Download
Cu111-C9H10aHtL.eps	54.77 kB	Download
Cu111-C9H10aHtL.jpg	154.49 kB	Download
Cu111-C9H10aHtL.eps	54.78 kB	Download
Cu111-C9H10aHtL.jpg	159.73 kB	Download
Cu111-C9H10trans.eps	54.78 kB	Download
Cu111-C9H10trans.jpg	160.98 kB	Download

# File sizes and upload options and examples

## File formats:

- All formats accepted

## Per file size limitations\*:

- 2.5GB
- Larger sizes with permission
- Max 1000 files per upload

## Tabular Ingest\*:

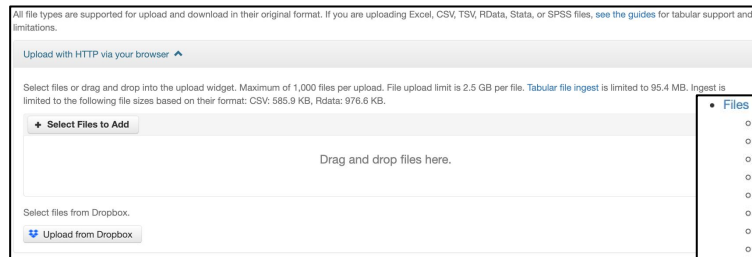
- 95.4 MB
- CSV 585.9KB
- Rdata 976.6KB

## Total project size limitations\*:

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

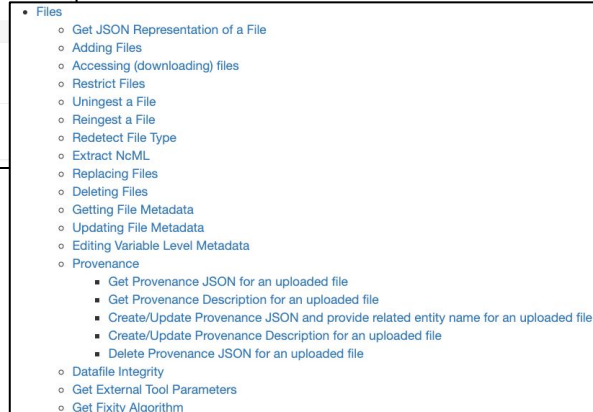
## Policies\*:

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



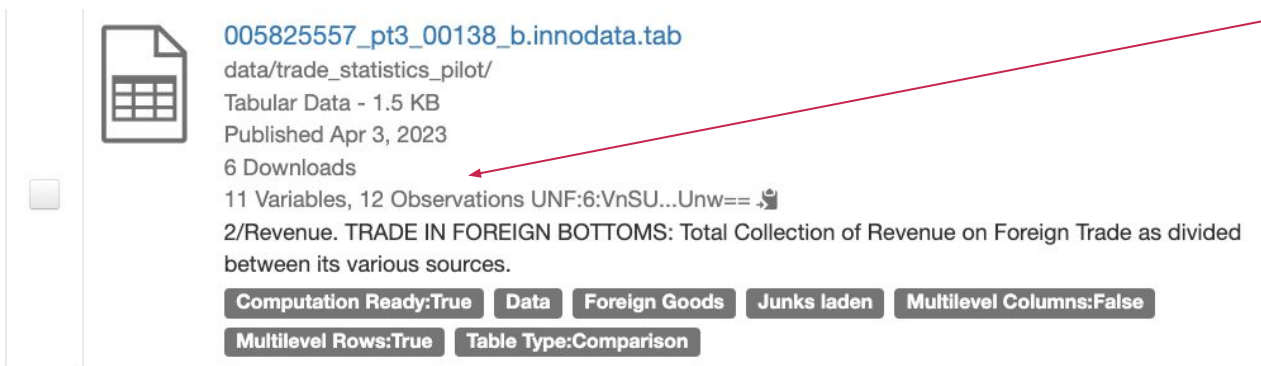
## File upload options

## Native APIs file handling options



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

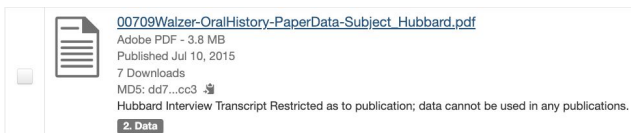
# File level metadata considerations and features



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

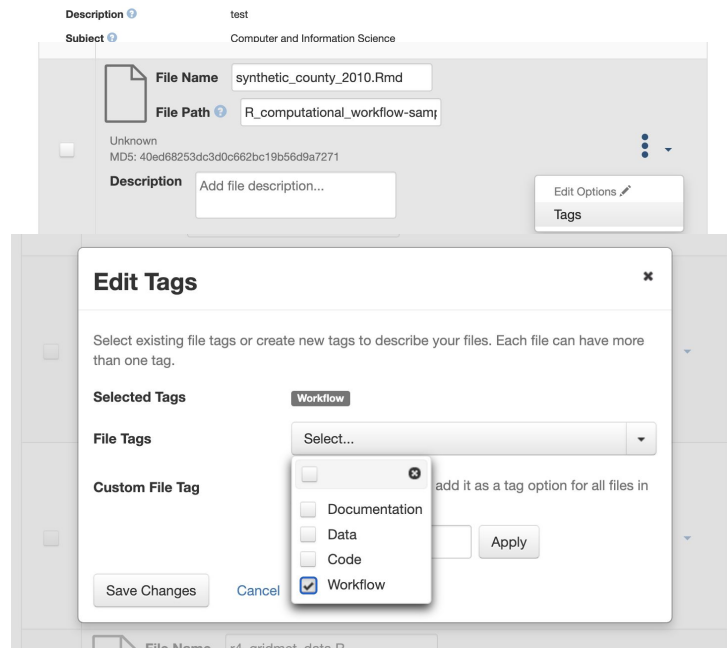
Customs Data JAMES H. HART STATISTICS

File level “tags”

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

# 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)



# 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](#).

**Access Dataset -**

Contact Owner Share

Dataset Metrics 1,797 Downloads

**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.](#)

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

Access File [v](#)

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

File Metrics [?](#)


50 Downloads [?](#)

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

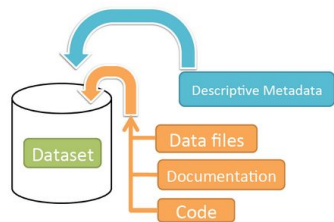
LLM kicks in



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

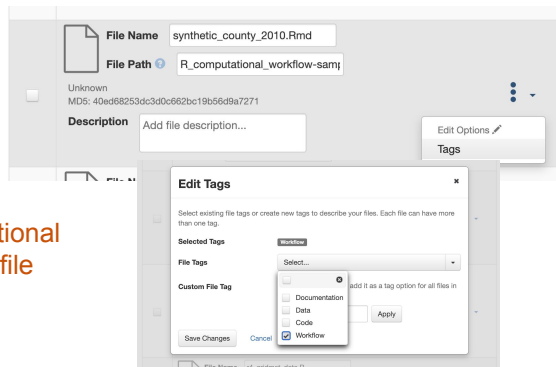
# File Types, Format, Documentation

Schematic Diagram of a Dataset in Dataverse 4.0

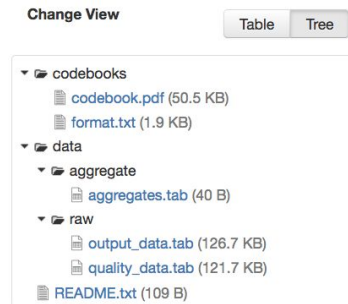


Container for your data, documentation, and code.

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

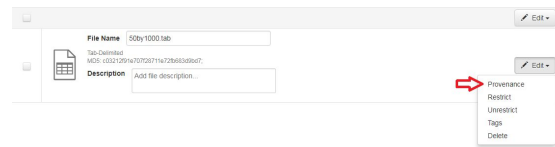
Public (1,784,339)

Restricted (50,591)

Embargoed then Public (154)

Embargoed then Restricted (14)

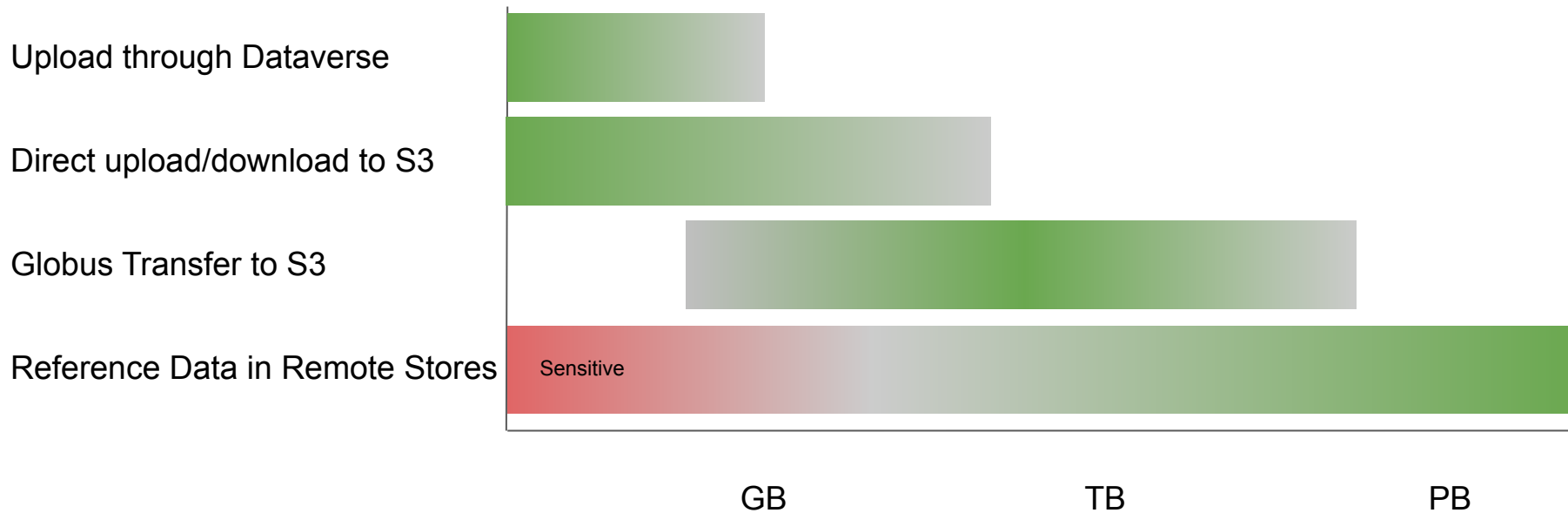
File level Provenance support



**Coming soon...**

# Big data support in Dataverse

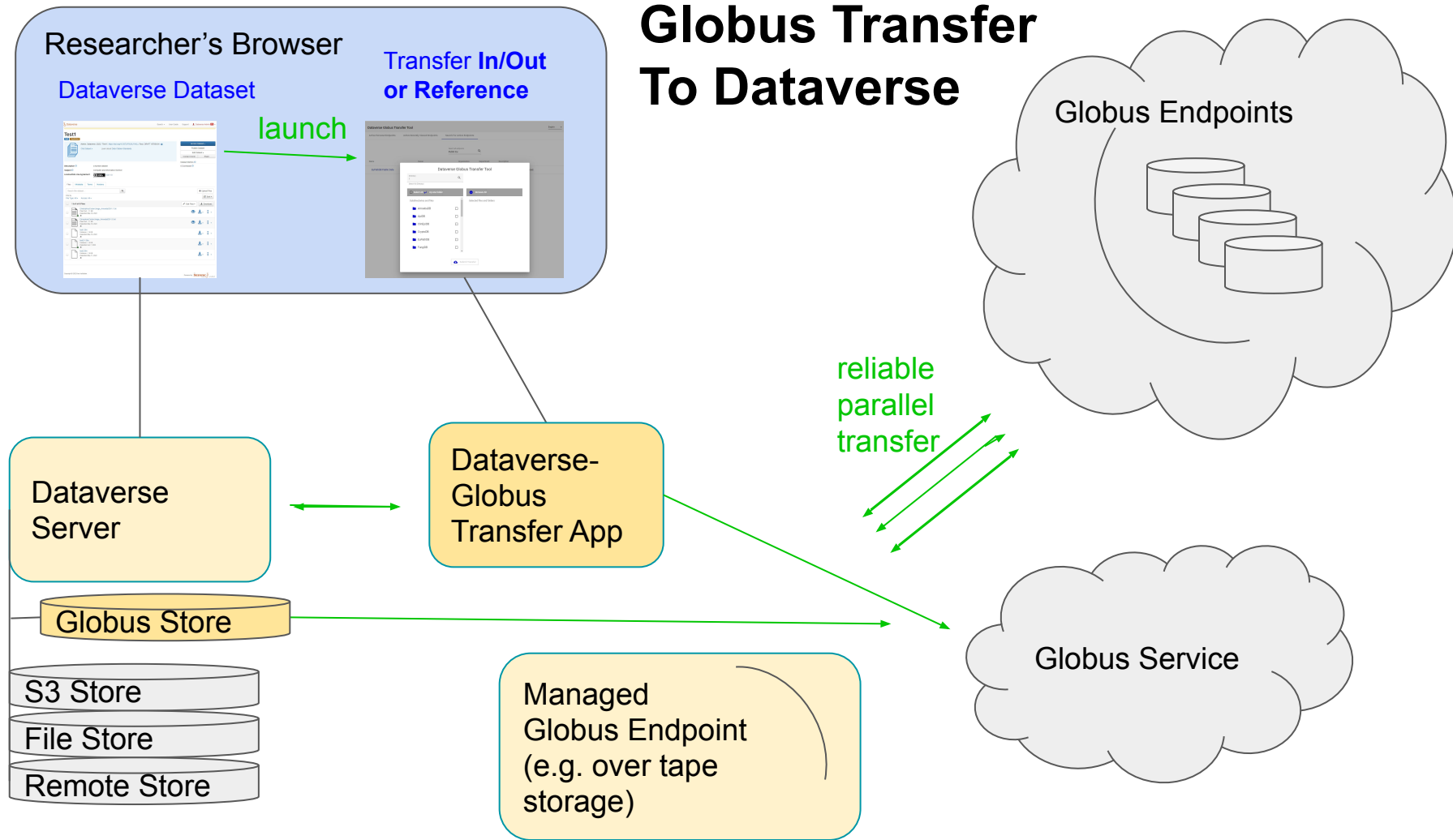
(overview from 2022)



We are now moving to use Globus API (rather than S3) for store interactions



# Globus Transfer To Dataverse





# Considerations

- Storage costs
- Initial data transfer costs
- File access costs
- File size limitations
- File count limitations
- Access protocol
- Backups
- User-data and application-data interaction requirements
- Dataset representation
- Usage monitoring
- Admin overhead

# 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.

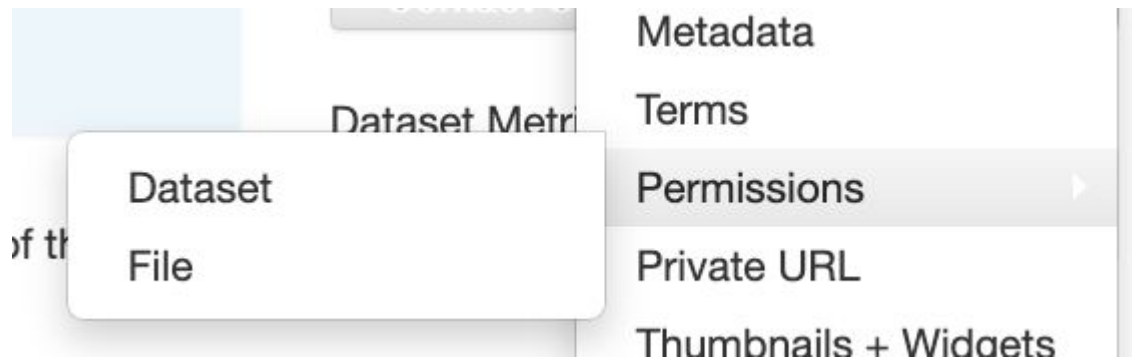
☒ Include Deleted Files

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
- Private URL for DRAFT dataset

## 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

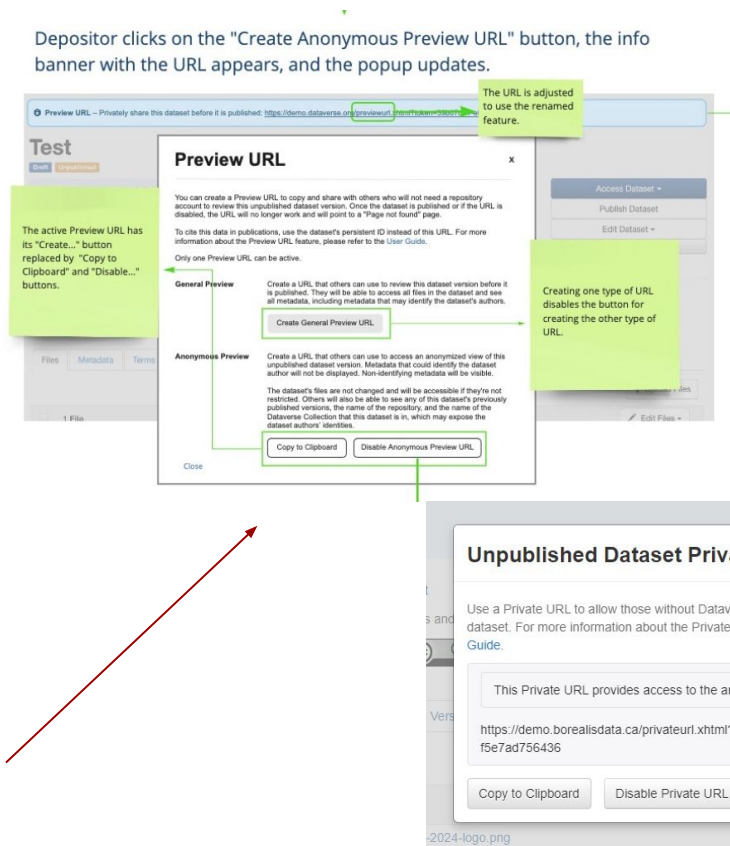
**Delete Dataset**

# 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
- Private URL for DRAFT dataset
- Anonymous Peer Review





# Enhanced File Handling of Data Types

# File Handling Checklist

Enhanced file handling and data management features include:

- Deposit & sharing (file identification, checksum (MD5), packaging)
- File Preservation (normalization)
- Tabular data ingest (tabular data formats)
- Geospatial and dimensional data (GeoJSON, Shapefile, GeoTIFF, NetCDF & HDF5)
- Additional file metadata extraction (FITS, NetCDF (NcML) and others)
- File previewers & data analysis integrations

Dataverse Format Checklist	Deposit & Sharing (Identification, MD5 checksum)	Preservation (Normalization)	Exploration (Preview and extraction)	Analysis (Tools & integrations)
CSV	Y	Y	Y	Y
TXT	Y	Y	Y	Y
MP4	Y	N	Y	Y
TIFF	Y	N	N	N
Quicktime (.MOV)	Y	N	Y	Y
Excel (XLSX)	Y	Y	Y	Y
SPSS	Y	Y	Y	Y
STATA	Y	Y	Y	Y
SAS	Y	N	N	N
Shapefile	Y	N	Y	Y
GeoTIFF	Y (beta)	N	Y (beta)	Y

# File Preservation

- Format identification (MIME Type)
- Checksum (MD5)
- Format normalization
- File metadata
- Folder preservation
- Exporting options for preservation system integration (BagIt, Archivematica)



[Green\\_Plains\\_2021b\\_inv.xyz](#)

Co-Ordinate Animation - 16.2 KB

Published Oct 4, 2023

0 Downloads

MD5: bee...e82



# 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>.

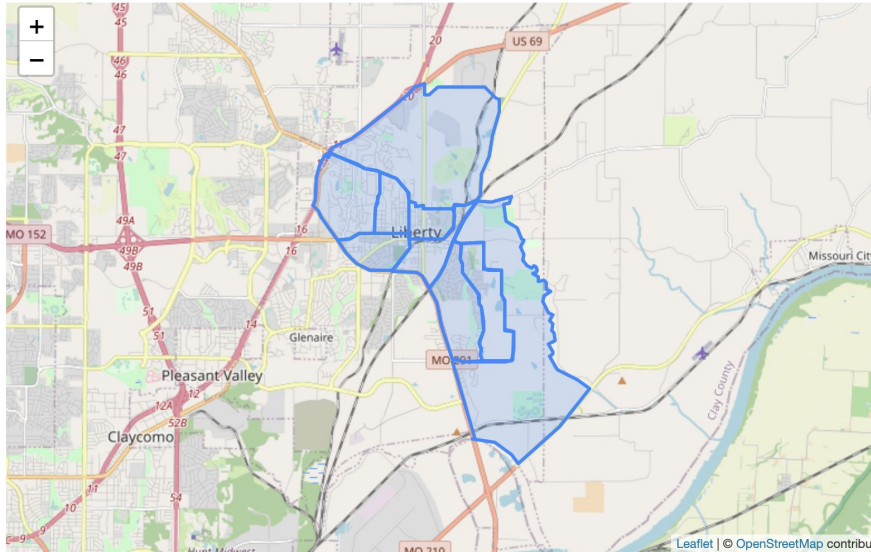




# GeoJSON Previewer

Preview Metadata Versions

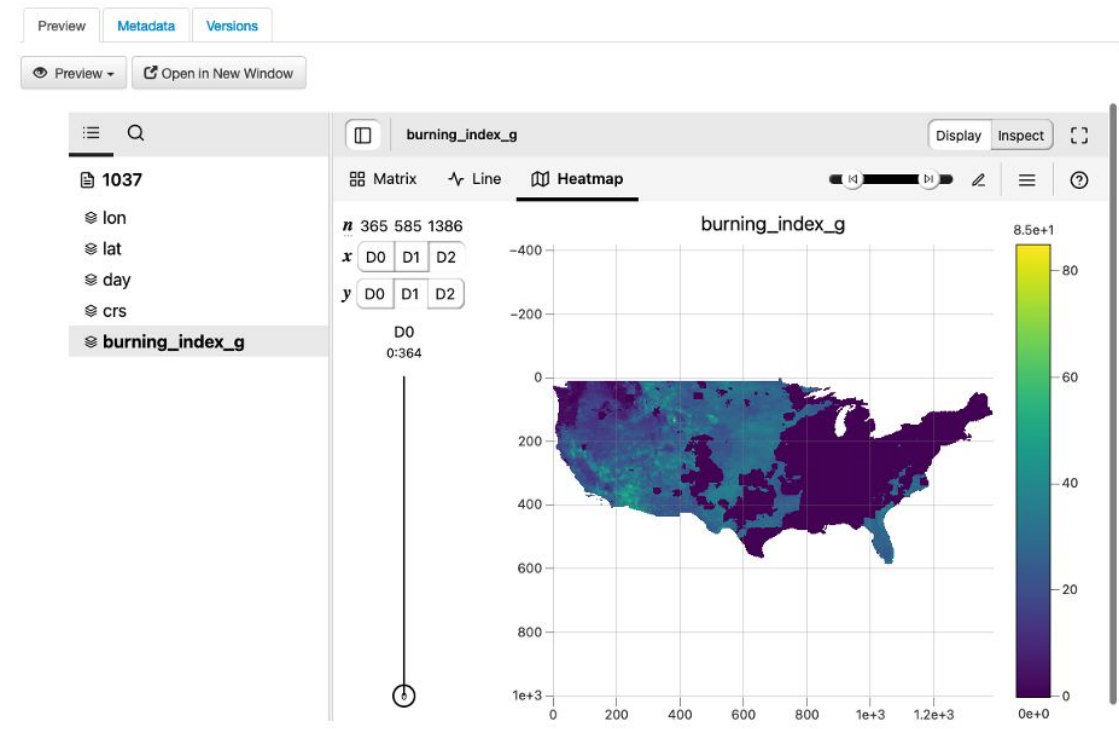
Open in New Window



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

# NetCDF & HDF5 Previewer

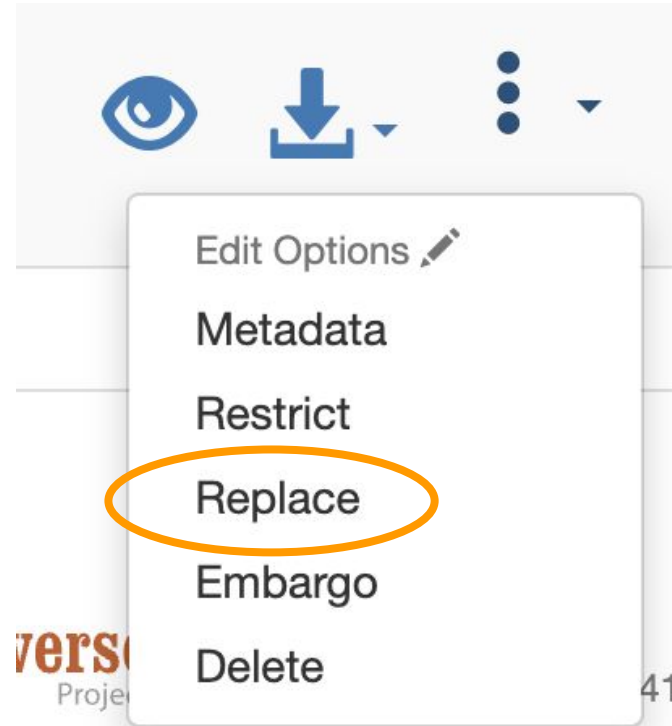
- Common scientific data formats
- Metadata extracted in NcML (XML) format and saved as an auxiliary file
- H5Web visualization and file preview tool
  - Preview data
  - Explore and visualize data



# **File Versioning**

# Versioning

- A dataset begins as an unpublished DRAFT; when published becomes v1.0
- Future versions can be minor (e.g. v1.1) or major, (e.g v2.0)
- File metadata can be changed per version, as needed
- File “versioning” is provided via file replacement - when uploading a new version of the file, it is linked to the previous version



# Persistent Identifiers

- PIDs are assigned for every dataset
- PIDs can also be assigned per file, configurable per installation
  - Can also be configured for only specific Dataverse collections within the installation
  - Can be defined to be “dependent” on the dataset PID or “independent”

## File Citation

Admin, Dataverse, 2023, "argentina.jpeg", *GPD Previewers*, <https://doi.org/10.70122/FK2/BLXVAF/WKNVG3>, Demo Dataverse, V3

[Cite Data File](#) ▼

Learn about [Data Citation Standards](#).

## Dataset Citation

Admin, Dataverse, 2023, "GPD Previewers", <https://doi.org/10.70122/FK2/BLXVAF>, Demo Dataverse, V3

[Cite Dataset](#) ▼

Learn about [Data Citation Standards](#).

# **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](#)

# External Tools Manifest

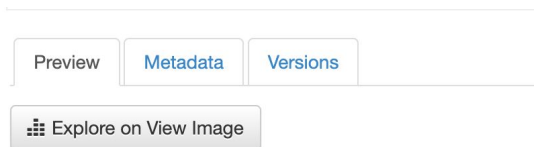
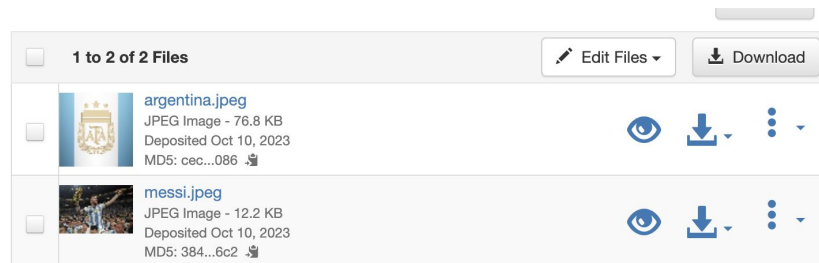
- External tools must be expressed in an external tool manifest file
- Can be uploaded to a Dataverse installation via API

```
{
  "displayName": "Awesome Tool",
  "description": "The most awesome tool.",
  "type": "explore",
  "toolUrl": "https://awesometool.com",
  "toolParameters": {
    "queryParameters": [
      {
        "fileid": "{fileId}"
      },
      {
        "key": "{apiToken}"
      }
    ]
  }
}
```




# 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



# 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 and downloader too

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

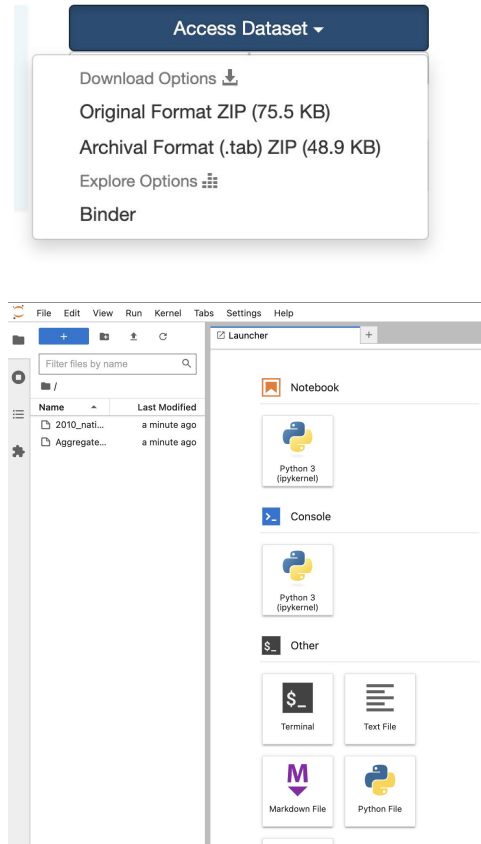
Add Group + Search Items per page 25 1 - 21 of 21 < >

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

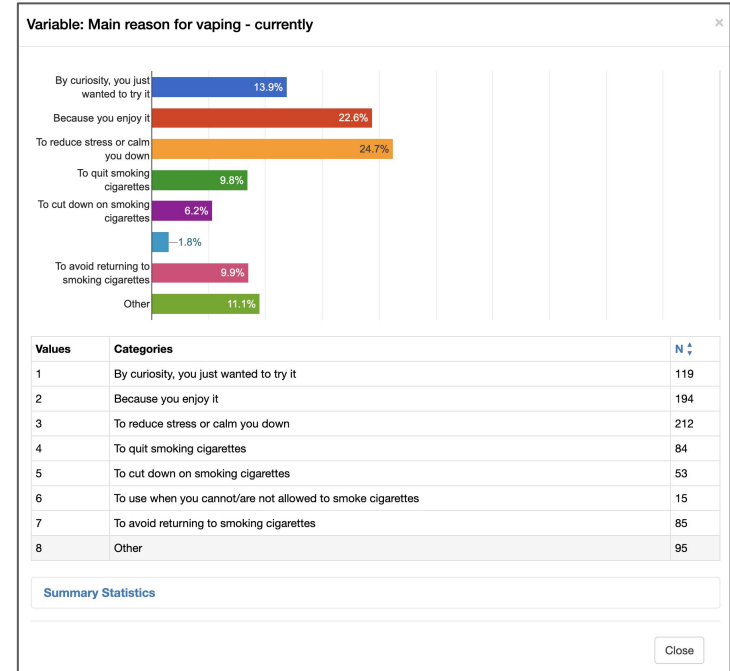


# Spotlight: Survey Data Analysis using Data Explorer

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

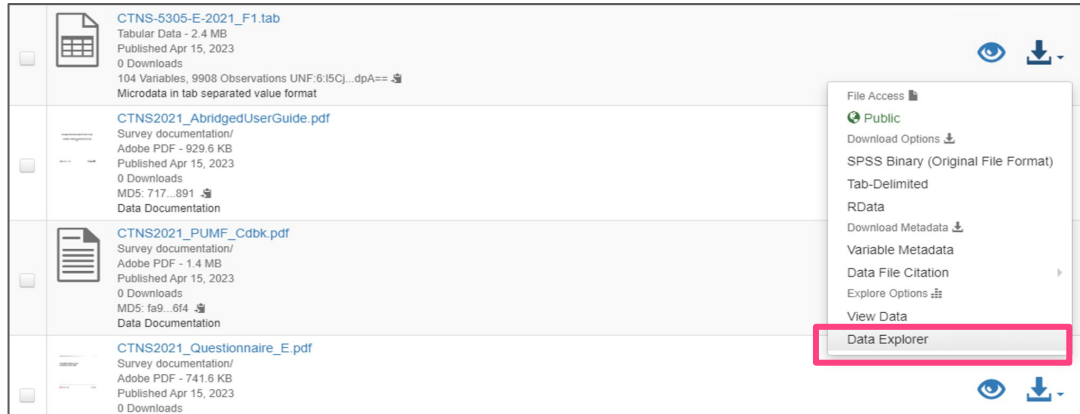
# Data Analysis using Data Explorer

- 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:QUEzj6HFqsjnX706WmLdQ== [fileUNF]

< Hide Groups

English

Cross Tabulation Summary Statistics Download

Items per page: 25 1 - 25 of 65

All Variables

Search

Confirm Type

Changes in Income

Parents

Benefits

Unemployment

Weight Variable

Sociodemographics

Additional Sources of Funding

Job Information

Parental benefits for spouses

	ID	Name	Label	Weight	View Summary Statistics	View Categories	View Questions
<input type="checkbox"/>	v6012811	PUMFID	Sequential Identification Number	v6012809			
<input type="checkbox"/>	v6012804	SRYR	Survey reference year	v6012809			
<input type="checkbox"/>	v6012769	REGION6	Province or region				
<input type="checkbox"/>	v6012819	SEX	Sex of respondent				
<input type="checkbox"/>	v6012801	AGECAT	Age of respondent - grouped				
<input type="checkbox"/>	v6012797	EDUC	Highest educational attainment				
<input type="checkbox"/>	v6012779	TYPES	Survey type as determined for the EICS	v6012809			
<input type="checkbox"/>	v6012822	WORKNOW	Respondent currently working at a job or business	v6012809			

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

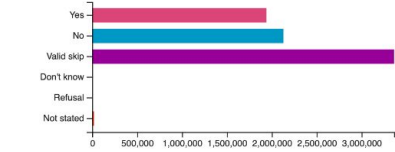
Job Information

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

Items per page 25 1 - 25 of 65

View Summary Statistics View Categories View Questions

Weight

v6012809

v6012809

v6012809

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	6020

# 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. Odesi)

## Data catalogs

The screenshot displays the Odesi data catalog interface. At the top, there's a header with the Odesi logo and a language selector set to 'Français'. Navigation links include 'My List (0)', 'Browse', 'Contact', and 'Deposit Data'. The search bar contains the term 'stress', and a dropdown menu shows 'Variable'. Below the search bar, there are filters for 'Collection' (set to 'All collections selected'), 'Sort' (set to 'Relevance'), and 'Date (YYYY)' (with 'From' and 'To' fields). A blue 'SEARCH' button is at the bottom of the filter section. To the right of the search bar, 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.

Results 1 to 20 of 513  
Page 1 of 26

Limit your search:

Collection

- Odesi 291
- DU 235
- Dlimt 184
- POP 52
- Borealis 38
- + More

Series

- Canadian Tobacco Use Monitoring Survey 40
- Enquête De Surveillance De L'usage Du Tabac Au Canada 34
- Canadian Community Health Survey 24

[Enquête sociale générale 2010, Cycle 24 \[Canada\]: Bien-être et stress lié au manque de temps, Fichier principal](#)

En pensant au niveau de **stress** dans votre vie, diriez-vous q En pensant au niveau de **stress** dans votre vie, diriez-vous que la plupart de vos journées sont : Quelle est votre principale source de **stress** ? Quelle est votre principale source de **stress**?

● Matching Variables 2

[General Social Survey, Cycle 24, 2010 \[Canada\]: Time-Stress and Well-Being, Main File \[version 5\]](#)

...being pressed for time was a positive stimulant as well as times the respondent felt **stress** because of the number of activities he/she had to do in a day. Feel constantly **stressed** trying to accomplish more than can handle?...you're constantly under **stress** trying to accomplish more

# Thank you and resources!

The Dataverse Project, IQSS, Harvard University:

<https://dataverse.org/> and [Guides](#)

The Harvard Dataverse Repository support and policies website:

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

Upcoming Trainings, **CHECK FOR UPCOMING WEBINARS!**

<https://support.dataverse.harvard.edu/calendar/upcoming>

Contact us:

[support@dataverse.harvard.edu](mailto:support@dataverse.harvard.edu)