Dataverse for data sharing: An introduction

Chantal Ripp & Jane Fry
November 3, 2021

Agenda

- What is Dataverse
- Benefits of Dataverse
- Using Dataverse for your research data
- Resources - many valuable ones!

Information about Dataverse in the following slides is taken from:

https://learn.scholarsportal.info/all-guides/dataverse/

What is Dataverse?

- A research data platform for sharing, discovering and preserving research data
- A repository for datasets
- In partnership with ScholarsPortal
  - a service of the Ontario Council of University Libraries (OCUL)
  - For subscribing institutions
    - Students, researchers, faculty and staff

Key benefits

- Robust Canadian storage network
  - Hosted in University of Toronto
    - On Ontario Library Research Cloud (OLRC)
      - A cloud storage network
      - Data is automatically replicated
      - Long-term access and preservation
- Data management
  - Meets one of the requirements of the Tri-Agency Research Data Management Policy
    - Data deposit
    - Data dissemination
    - Open access (when permitted)
- Access control
  - You determine who has access to your data
  - IP controlled

More key benefits

- DOIs and citation standards support
  - DOIs - digital object identifiers
    - International standard for simple, persistent identification for access to your data on the web
    - Automatically provided
  - Some journals are requiring DOIs for datasets associated with a journal article
    - This is often just a replication dataset
    - Mandatory for major Economics journals
    - Other journals will be following soon
  - Standard and rich metadata fields
    - Allow for maximum discoverability for your datasets
  - Dataverse automatically generates a data citation
    - Citation can be downloaded in various formats
    - Can be used in any research publication
    - Simply click the “Cite Data” button

More ...

- Discoverability, recognition and increased visibility
  - Share your data with a global research community
  - Researchers will get recognition and proper academic credit for their scholarly work
  - Indexed in
    - re3data.org
      - Registry of research data repositories
      - More than 2450 research data repositories
      - Most comprehensive source of reference for research data infrastructures globally
    - DataCite's Global Search Service
    - Web of Sciences' Data Citation index
  - Dataverse is regularly crawled by Google
  - *Note: only published and public files have their metadata harvested*

What File Size?

- Dataverse supports file sizes less than 3GB
  - If larger, you can zip the files
  - If it is a tabular file, you can break it into smaller files (with an accompanying *Readme file*)
- Recommended to keep files to less than 500MB for tabular data files that require more time to upload
  - For file sizes larger than 500MB, contact Dataverse support
- No limit on total size of all files you can upload for a single dataset
  - If more than 10GB in total file size, contact Dataverse support
- See Scholars Portal Dataverse Guide for more info
- If you are working with big data
  - Check out FRDR (Federated Research Data Repository)

What types of files?

- Any file type can be published
  - Shapefiles
  - Images
  - Compressed files
  - Stata
  - SPSS
  - Rdata
  - .wav
  - .jpeg
  - .png
  - Audio files
  - GIS data
  - Csv
  - Tab
  - Ascii
  - ...

Collaboration

- Can I collaborate with colleagues in a Dataverse?
  - Yes! Has to be hosted with a subscribing institution.
  - Dataverse supports groups of researchers working on the same research project.
  - Very useful when working with researchers at other institutions.
  - Can assign different group permissions.
  - Can assign different account roles.
  - Can also share your dataset without publishing it.

Source: https://blog.jostle.me/blog/why-collaboration-is-important

Safety of your data

- Scholars Portal has backup copies
- Can be restricted to authorized users only
  - Individual accounts
  - Specific IP groups, eg., your institution only
- Can keep a copy of your research data in Dataverse
  - As a backup copy that you are keeping elsewhere

Dataverse vs Dataset

- A Dataverse Collection
  - Structure is much like the structure of Windows Explorer
    - main folders/dataverses
      - sub-folders/sub-Dataverses
  - It is a folder/bucket that can have other Dataverses in it
  - Can also have datasets in it
- A dataset
  - must be in a Dataverse
  - contains data files and metadata about those data files

Dataverse vs a Website

	Dataverse	*Website*
Sustainability	No worries	Keep an eye on it (not ‘one and done’)
Discoverability	Maximum	Need proper SEO (Search engine optimization)
Security	Infrastructure designed to protect against system malfunction, malicious attack, or other technical issues	Hopefully won’t be hacked, but ...

- Can use both together
  - Put your dataset in Dataverse and the DOI on your website
  - Then your research project can have its own branding

More FAQs

- Can I test it out to see how it works?
  - Sure!
  - https://demodv.scholarsportal.info/
  - Try it out to see if you like its features
- What is the cost to use Dataverse for the researcher?
  - Generally, no cost for the researcher
  - If your institution subscribes to Scholars Portal Dataverse
    - No direct cost to researchers
    - Must use your institutional email address
  - Check with your local Data Support Services to find more information

Still more!

- How much data can I download from Dataverse?
  - No restrictions
  - As much openly available data as you want
- Can I upload .zip bundles?
  - Yes!
  - Dataverse will automatically unpack your single zip package and present individual files
- How many files can be in my .zip bundle?
  - Less than 500 because of additional processing that is required if there are more

Dataverse limitations

- Dataverse may not be suitable if your data ...
  - Requires
    - high volume computational processing infrastructure to share or access
    - an automatic embargoed period
      - You would have to manually embargo it
  - Is sensitive data
    - Dataverse cannot accept any data that contains confidential or sensitive information

Making your data FAIR

- Good research data practice
  - data management and stewardship
- Findable
  - discoverable
- Accessible
  - Data can be downloaded
- Interoperable
  - Data can be integrated with other data and can be analyzed
- Reusable
  - Metadata and data are well-described
  - They can be replicated and/or combined in different settings
- Here's a checklist to follow:
  - FAIR-Aware
  - FAIR-Aware Tool video (1.21 min)

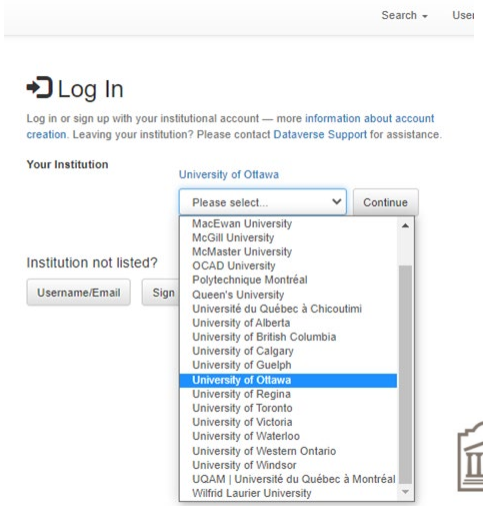
Time for a Demonstration!

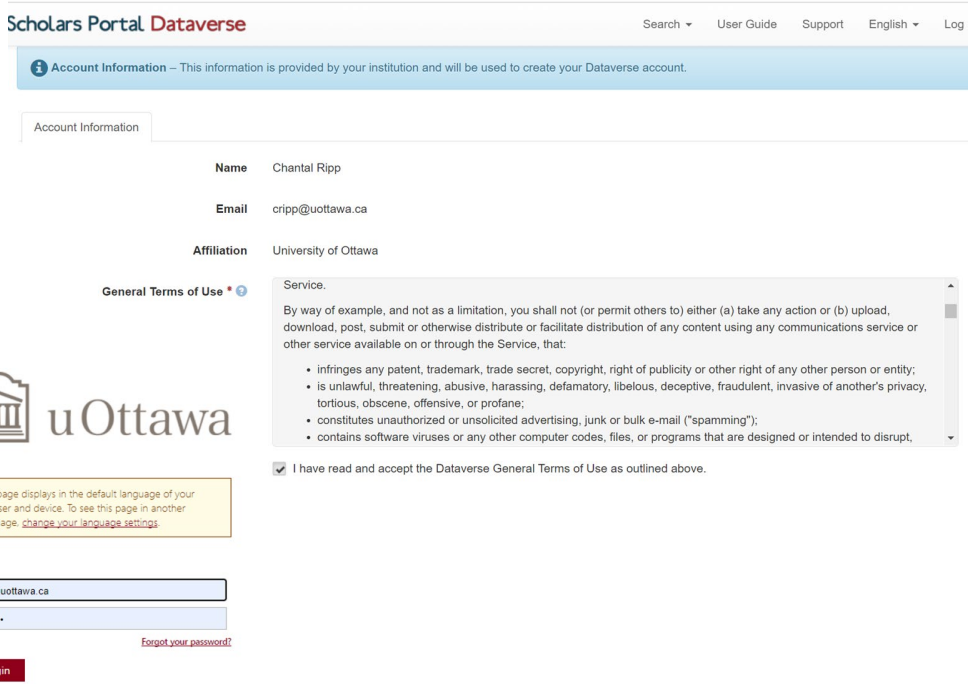
Creating Your Account

Option 1

Institutional Single Sign on

- sign-in using the same username and password you use to access your institution's online services



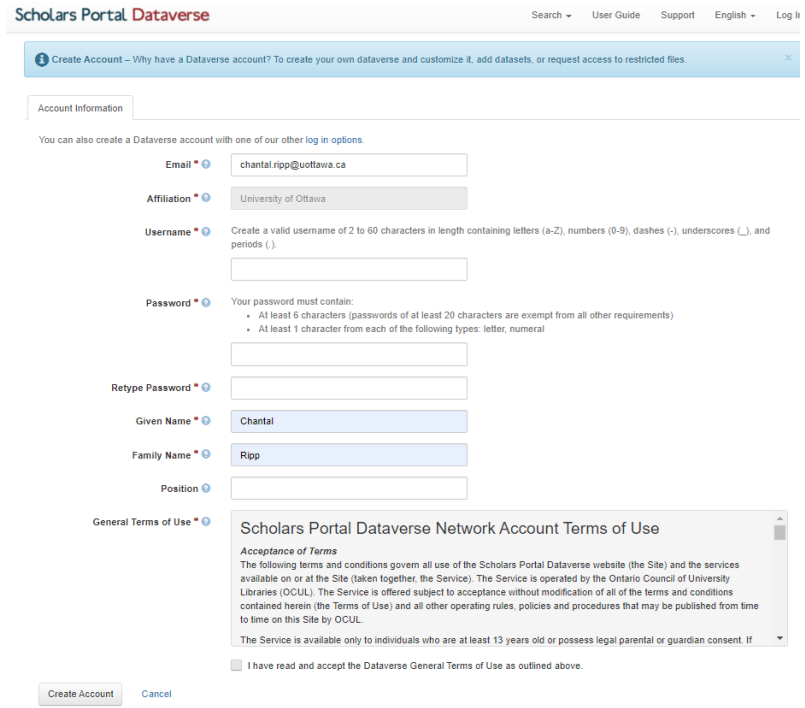


Creating Your Account

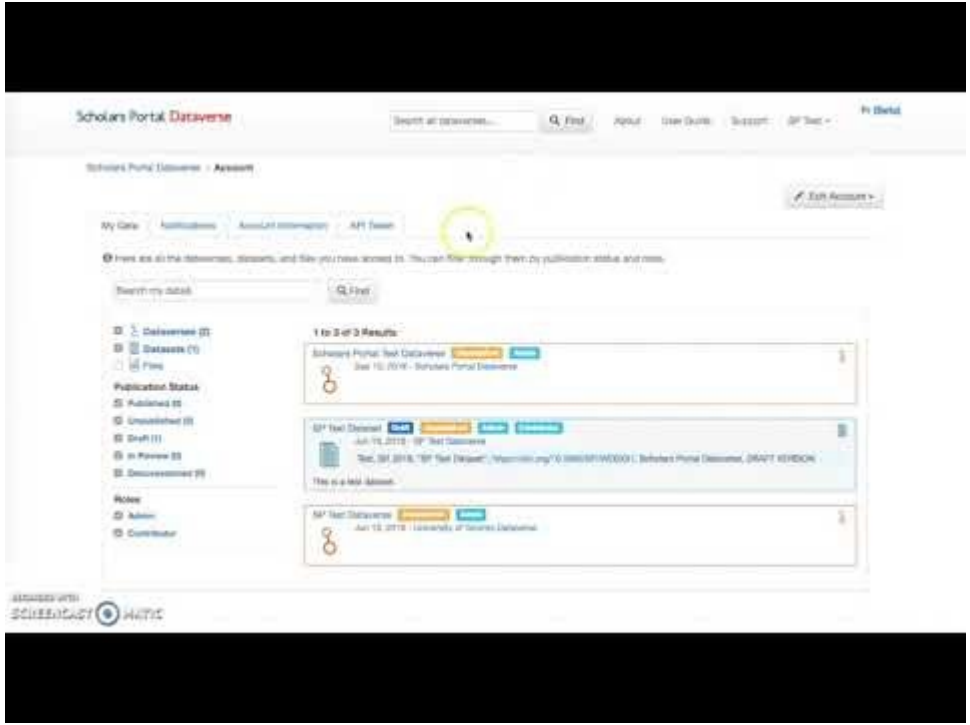
Option 2

Username/password

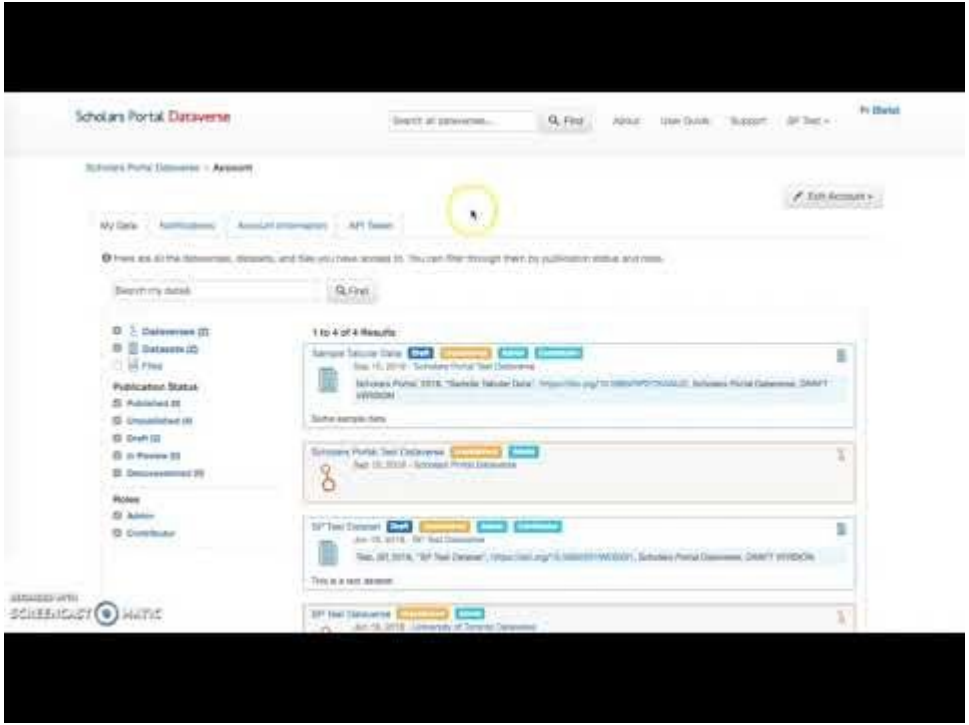
- If you are affiliated with an institution, please use your institutional email address. Users that are not affiliated with a participating institution will have limited access.
- Refer to SP FAQ for more information.



Creating, Editing, and Publishing a Dataverse



Publishing Datasets in Dataverse



Modifying Terms

Want to apply a different licence? Click “Edit Terms Requirements”

Add another open licence under “Terms of Use”

Dataverse and FAIR

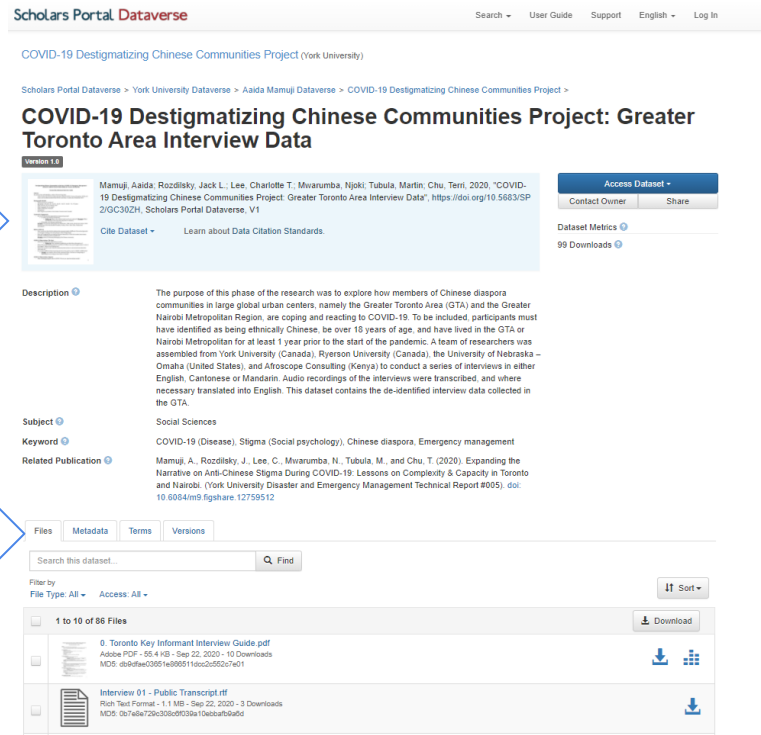
Data citation with persistent identifier (DOI)

Data files

Metadata

Terms/Data Licence

Dataset version



Adapted from Brooke, Danny. (2019, May). Community Built Research Data Infrastructure: The Dataverse Project. Presented at the IASSIST 2019: Data down under: Exploring "data firsts", Sydney, Australia: Zenodo. http://doi.org/10.5281/zenodo.3612399

Resources

- Guide to depositing data in uOttawa Dataverse
- Carleton University Dataverse
- Scholars Portal Dataverse
  - Dataverse overview: Finding and exploring data (7 min video)
  - Creating, editing and Publishing a Dataverse (6 min video)
  - Publishing datasets in Dataverse (6 min video)
- Digital Research Alliance (former Portage)
  - Can I share my data?
  - Documentation and supporting material required for deposit
  - Recommended repositories for COVID- 19 research data

More resources

- Dataverse North metadata best practices guide
- Introduction to the Harvard Dataverse Repository (1 hr 36 min video)
- FAIR
  - FAIR- Aware Tool video (1.21 min)
  - FAIR- Aware Tool
  - How FAIR are your data?

Thank You!

Contact Information

Chantal Ripp

Research Librarian (Data) |

Bibliothécaire de recherche (Données)

309E, Bibliothèque Morisset Library

University of Ottawa | Université d'Ottawa

chantal.ripp@uottawa.ca

Jane Fry

Data Services Librarian

MacOdrum Library

Carleton University, Ottawa

jane.fry@carleton.ca