Dataverse for data sharing: An introduction

Chantal Ripp & Jane Fry
June 16, 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/faq/
What is Dataverse?

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

- Robust Canadian storage network
  - Hosted in University of Toronto
    - On Canadian servers
  - 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
    - An 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
    - Can be used in any research publication
    - Citation can be downloaded in various formats
    - Simply click the “Cite Data” button
● 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
      ● The 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 of UoO/Carleton
● 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 of UoO/Carleton
● 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
  - ...

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](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 IPs
  - 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
  ○ 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

<table>
<thead>
<tr>
<th></th>
<th>Dataverse</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sustainability</strong></td>
<td>No worries</td>
<td>Keep an eye on it (not ‘one and done’)</td>
</tr>
<tr>
<td><strong>Discoverability</strong></td>
<td>Maximum</td>
<td>Need proper SEO (Search engine optimization)</td>
</tr>
<tr>
<td><strong>Security</strong></td>
<td>Can’t be hacked</td>
<td>Hopefully won’t be hacked, but ...</td>
</tr>
</tbody>
</table>

- 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/dataverse/sp?q=demo+dataverse](https://demodv.scholarsportal.info/dataverse/sp?q=demo+dataverse)
  - Try it out to see if you like its features

- What is the cost to use Dataverse?
  - None!
  - University of Ottawa and Carleton belong to the consortium that supports it (OCUL)
    - Therefore, no direct cost to researchers
    - Must use your institutional email address
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)
Creating Your Account

Options 1

Institutional Single Sign on

- sign-in using the same username and password you use to access your institution’s online services
Options 2

Username/password

- If you are affiliated with an institution, please use your institutional e-mail 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

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)
- **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!

Source: https://workingnation.com/2020-presidential-debate-future-of-work-questions/
Contact Information

Chantal Ripp
Research Data Management Librarian (interim) |
Bibliothécaire de gestion des données de recherche (intérim)
309E, Bibliothèque Morisset Library
University of Ottawa | Université d’Ottawa
chantal.ripp@uottawa.ca

Jane Fry
Data Services Librarian
MacOdrum Library, Carleton University
jane.fry@carleton.ca
https://library.carleton.ca/find/data