## Depositing and Managing Data

### Supported File Formats

Any kind of file can be uploaded to Dataverse, but extra functionality is supported for some filetypes:

**Tabular data**

Tabular data (Stata, SPSS, Excel, R, & CSV) is normalized to .tab format on upload—a non-proprietary archival format that a variety of programs can read. See [bit.ly/2VmkTGF](https://bit.ly/2VmkTGF) for more about this process. Normalization is important for long-term preservation of digital data; however, the deposited files can be downloaded in multiple formats, always including the original. You will choose the format from a list when you click Download.

In Dataverse, tabular normalization also allows you to perform statistical data exploration and visualization right in your browser. Click the Explore button in the Files tab to see what it can do! Learn more about the TwoRavens Data Explorer application at [bit.ly/2H2QQQC](https://bit.ly/2H2QQQC).

**Compressed Files (.zip, .tar)**

Compressed files are the preferred method for uploading large datasets or many files to Dataverse.

Zip files are automatically extracted on upload, and the contents will appear as a list under the Files tab. Folder structure and file hierarchy within the zip file are maintained on extraction (as of DV version 4.11).

**Geospatial shapefiles** and **Astronomy data (FITS)** also have special functionality in Dataverse. Contact [research.data@ubc.ca](mailto:research.data@ubc.ca) for more information if your data involves these filetypes.

### File Size

Dataverse is not intended to handle Big Data. Current file size limits are:

- **For upload:** 2.5 Gb per file (unlimited number of files in a dataset)
- **For tabular normalization:** 500 Mb per file (tabular files over this size will remain in their original format, e.g. Excel)
- While there is no cap on the overall number of files that you can upload, if your data exceeds 10 Gb, please contact [research.data@ubc.ca](mailto:research.data@ubc.ca) to discuss the best repository options, as we have other solutions.

### Access Control

Sensitive files can be Restricted so they are not freely available to download. Files can be either fully restricted, or set up so users can send access requests to the contact email for review.

**To restrict a file:**

Under the dataset Files tab, click the checkbox beside the file(s) you want to restrict, then select Edit Files -> Restrict. In the popup window, you may describe the conditions of use or license for this file in the Terms of Access box. Check the Request Access box to allow users to send the contact person an email asking for permission to download a restricted file. Click Continue to finish.

Files can also be Unrestricted if the terms change—for example, once an embargo period has passed.

**To Unrestrict files** click the checkbox for the restricted file under the Files tab, then Edit Files -> Unrestrict.

### Licenses and Terms

You have control over how your data can be used. Dataverse allows for a variety of licenses and terms of use:

- **Built-in License Templates**—can be selected at dataset creation, and changed at any time. These automatically apply the right information about the license to the dataset's metadata.

Apply a template to your dataset by selecting one from the drop-down menu under Dataset Template, at the top of the New Dataset creation form.

- **Custom:** One-size-fits-all licenses don’t suit every dataset, you can create customized terms and conditions of use.

**To customize terms:**

Any dataset saved in Draft or Published form will have a Terms tab on the dataset page. Click **Edit Terms Requirements** and fill in the form with your Terms of Use and/or Terms of Access for Restricted files. If you need help creating a custom license, contact [research.data@ubc.ca](mailto:research.data@ubc.ca)

### Version control

Every change made to a dataset—adding files, editing metadata, etc.—is saved as a new version of the dataset. This allows you to track the change history of the project, which can be viewed under the Versions tab. This is useful if you need to roll back to a previous version, or find out who made what changes, when.
Using UBC Dataverse @ Scholars Portal

Sharing Data

**WHY DEPOSIT DATA?**
Storing your data in a repository like Dataverse makes it easy for other researchers, publishers, and collaborators to find, use, and cite your data. It supports the Open Science movement goals of transparency, accessibility, and collaboration, and can increase attention to your work.

Further, the Tri-Agencies in Canada (NSERC, CIHR, SSHRC) will soon be introducing policy requiring grant recipients to deposit their project data in a digital repository. You can review the draft policy at [bit.ly/2H3mGLv](https://bit.ly/2H3mGLv).

**Where can UBC Dataverse datasets be discovered?**
Data deposited in UBC Dataverse @ Scholars Portal is indexed by, and integrated with, many services on the Web, including:

- **DataCite**—This major global nonprofit organization provides DOIs, indexes the metadata for every object assigned a DOI, and offers a search interface for those objects at [https://search.datacite.org/](https://search.datacite.org/)
- **ORCID**—Adding your ORCID to your dataset metadata links it to your body of work, ensuring researchers can easily find all your publications.
- **FRDR**—Canada’s Federated Research Data Repository. FRDR provides a centralized platform for searching the contents of dozens of Canadian data repositories
- **Google/Google Scholar/Google Data**—Though data in Dataverse is not automatically searchable by Google Scholar, the UBC Dataverse has been customized so its deposited datasets are findable using this popular search tool.
- **Bing**
- **OpenAIRE** (coming soon!)

**Social media**
Spread the word about your research and improve your altmetrics by sharing your linked data on social media!

Dataverse provides a **Share** button for

1. each dataverse
2. each dataset

This button gives you the option to create a post with a link to your data on Facebook or Twitter.

**Linking to your dataset**
There are 2 main ways to direct people to your dataset:

- **DOI**—A permanent digital object identifier assigned to the dataset by Scholars Portal when the first Draft is created. Once the dataset is **Published**, anyone can use the DOI link to find it. While the dataset is still unpublished, the DOI can only be used by Dataverse account holders with permission to view that dataverse.
  - Use DOIs in citations, on your personal or research group website, in publications, and anywhere else you want to be sure the link to your data will remain stable over time
  - See [bit.ly/2VJtIHdM](https://bit.ly/2VJtIHdM) for more information about DOIs
- **Private URL**—A temporary link for use with **unpublished data**. The dataset can only be seen by those who have the link, and users do not need a Dataverse account. Great for giving pre-publication access to journals, reviewers, and collaborators.
  - To create a Private URL, on the dataset page, click **Edit -> Private URL -> Create Private URL**. The generated URL is saved in this location until the dataset is published, so you can copy it again and again as needed.

Questions? Contact [research.data@ubc.ca](mailto:research.data@ubc.ca)