

uOttawa Dataverse Deposit Guidelines

Library

biblio.uOttawa.ca



uOttawa

uOttawa Dataverse Deposit Guidelines

Version 1.0

July 14, 2021

The following guidelines cover preparing to deposit your data in uOttawa Dataverse, depositing your data and retrieving your own data or others'.

Preparing your data for deposit

Before depositing your data, make sure your dataset follows these best practices.

Sensitive data

In accordance with the [Scholars Portal terms of use](#), ensure that:

- Data are anonymized and sensitive or personally identifiable information is removed.
- If the dataset includes secondary data, you have the proper permissions to share the dataset.

File naming conventions

Following file naming best practices increases the discoverability of your dataset. Ensure that your dataset meets the following standards:

- Consistent file naming structure.
- Descriptive but short file names (fewer than 25 characters)
- No spaces between words: use underscores and capital letters instead (e.g., File_Name_20210701).
- Present dates in this format: YYYYMMDD
- No special characters, such as - @ # \$ % & * () +? > <

Preferred file formats

Dataverse accepts a wide range of formats. For long-term access and preservation, ensure that your data is in a preferred file format (such as those below). Also, make sure that all your files contain a valid extension (e.g., .txt, .pdf).

The following is a non-exhaustive list of preferred file formats for depositing in uOttawa Dataverse. If you have questions about preferred file formats, [email repository staff](#).

File Type	Preferred File Formats (examples)
Audio	<ul style="list-style-type: none"> • Uncompressed and lossless WAV or AIFF (.wav/.aiff) • Compressed and lossless FLAC (.flac) • Compressed and lossy MP3 (.mp3)
Image	<ul style="list-style-type: none"> • Uncompressed TIFF (.tif or .tiff) • Compressed and lossless PNG (.png) • Compressed and lossy JPEG (.jpg)
Text	<ul style="list-style-type: none"> • PDF (.pdf) • Tables (tab-separated Unicode plain text (.txt)) <p>If formatting needed:</p> <ul style="list-style-type: none"> • XML(.xml)
Markup language	<ul style="list-style-type: none"> • XML (.xml) • HTML (.html) • Related files: .css, .xslt, .js, .es

Transcription	<ul style="list-style-type: none"> • PDF (.pdf)
Video	<ul style="list-style-type: none"> • MPEG-4 (.mp4)
Statistical analysis	<ul style="list-style-type: none"> • R (.R, .RData) • SPSS (.dat/.sps) • STATA (.dat/.DO)

Metadata

Using rich metadata to describe your dataset helps support its discovery, reuse and preservation. Complete all required metadata fields and subject-specific metadata fields, as appropriate. For guidance, see the [Dataverse North Metadata Best Practices Guide](#).

Supplemental documentation

Supplemental documentation such as ReadMe files or codebooks enable your research data to be interpreted correctly. All deposited datasets must include a ReadMe or a codebook. The supplemental documentation should provide details on the files and their content including:

- The relationship between files
- Directory listing and file names, including content descriptions
- File formats
- File version information
- Explanation of any codes, acronyms and abbreviations used
- Procedures for data collection, analysis and processing
- Information about dataset completeness
- Limitations on reuse

ReadMe files must:

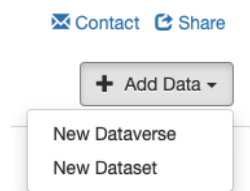
- Be saved as Unicode UTF-8 plain text files (.txt) or in PDF/A format (.pdf).
- Use forced numbering in the file name (e.g. 00_ReadMe.txt) to make it appear at the top of the file overview in Dataverse.

For a basic template, see the [Cornell University ReadMe template](#).

Depositing your data

Create a draft dataset

Once you are ready to deposit your data in the Dataverse, select the **Add Data** button on the right-hand side of the page and then **New Dataset**.



Enter metadata

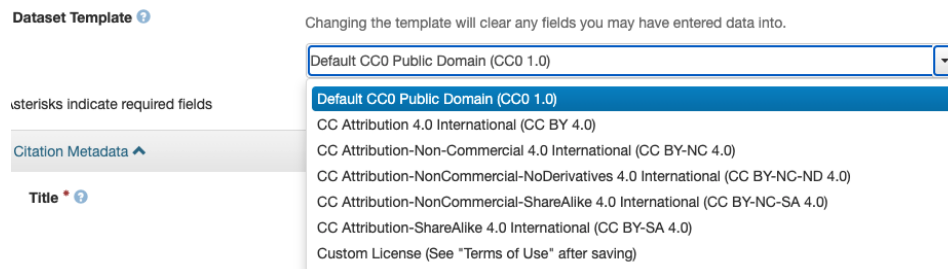
You can get information about the various metadata fields by placing your pointer on the field names. Some types of metadata you must provide include:

- **Title:** Enter a title for your dataset. If your dataset is used in a publication, you can enter the title of the publication and select **Add “Replication Data for” to Title**.
- **Author:** Enter your name as you use it in your publications. We recommend you add your affiliation as well. For entering co-authors, select the **plus button**. We also recommend you add your [ORCID](#).
- **Contact:** Enter a contact email address. Also add the name of the contact person or research group or institution.
- **Description:** If you need to add paragraphs, add the HTML tags `<p>` and `</p>` around each paragraph. If relevant, enter information about the data collection/methodology.
- **Keyword:** Enter information such as the subject area(s) (e.g., morphology or digital humanities) and the statistical method(s). Enter each keyword separately. To enter more keywords, select the **plus button**. **Vocabulary** and **Vocabulary URL** are not mandatory and can be left empty.
- **Related Publication:** If the files you are depositing are the background data for a publication, include a reference to the publication here.
- **Depositor and Deposit Date:** These fields are pre-filled. **Do not change them**. Deposit Date refers to the date when the dataset first was uploaded to the repository.

Please note: If your manuscript has been submitted for review but has not yet been accepted, **don't** list the name of the journal or publisher. Instead, just enter something like “Submitted for review.”

Specify data licence and attribution

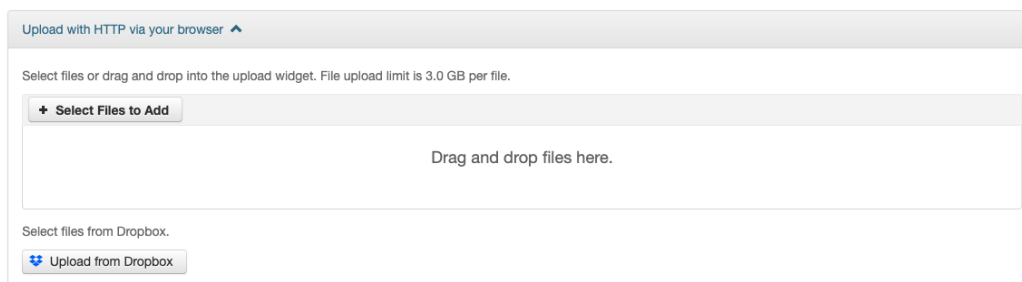
The default licence for reuse of data archived in uOttawa Dataverse is Creative Commons 0 (CC0). You can change this licence in the “Dataset Template” section.



If you're not sure which licence to use, [email repository staff](#).

Upload data files

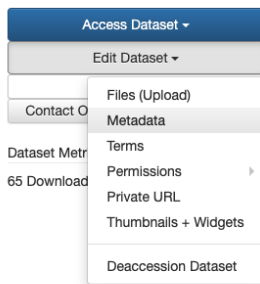
- Please note that uOttawa Dataverse has a file size limit of 2.5GB. However, there is no limit to the number of individual files in a folder. If your files are too large, try zipping your folder or [email repository staff](#) to discuss your options.
- To keep the folder structure of your dataset, pack the folders and single files into a container file (.zip).
- Select **Select Files to Add** and select your files from your computer or server space. You can select multiple files using the Shift + arrow keys or Ctrl + mouse click. You can also upload files from Dropbox or drag and drop your files in the designated area.
- When you have uploaded all the files in your dataset, select **Save Dataset**. A draft of your dataset will be saved.



Enter more metadata

After creating a dataset draft, you should enter more information about the dataset. This will increase the chances of others being able to discover your data and interpret and

reuse it correctly. Select the **Edit Dataset** button on the upper right-hand side of your dataset page.



If possible, fill out the following additional fields:

- **Language:** Select the language you have used to describe your data.
- **Contributor:** Credit those who have contributed to the dataset, including those who won't be listed as authors. In the **Type** list, you can choose what role they have had (e.g., Data Collector).
- **Grant Information:** Use the funder's full name.
- **Time Period Covered:** The time period the data is from or about.
- **Date of Collection:** When the data was collected or generated.
- **Kind of Data:** Hover your mouse over the question mark to the right of the field name, and you will see some suggestions about data types (e.g., survey data, experimental data, observation data).
- **Related Material:** Material related to the dataset.
- **Related Datasets:** Other datasets related to the current dataset. This can be your own or others' datasets. If available, use a full reference, including the persistent identifier (e.g., DOI).
- **Data Sources:** If you have not generated or collected the data yourself, enter information here about your sources. This can include an archive, a corpus or a website from which you downloaded the data.

File embargo

You can restrict access to (some of) your files for a period. During this embargo period the selected files will not be accessible, but the metadata about your dataset will be visible. Please note that restrictions on files need to be removed manually in Dataverse. To restrict access to a file, follow these steps:

- Upload the file (see "Upload data files" above).
- Select the box to the left of the name of the file you want.
- Select the **Edit files** button above the file section to the right, and then **Restrict**.
- When you have restricted all necessary files, select **Save Dataset**, and enter the terms of access in the pop-up window. If you want users to be able to contact you to request access, select the **Request Access** box. Select **Continue**.

- Select the boxes to the left of the file once again.
- Select the **Edit Files** button above the file section to the right, and then **Tags**. In the **Custom File Tag** box, type “Not available until YYYY-MM-DD.” Select **Apply** and **Save Changes**.
- Specify the date when your file(s) will be made accessible using the metadata field **Distribution Date** (see “Enter more metadata” above).

Publish your dataset

- Once you have entered all the necessary metadata and uploaded your files, select **Submit for Review**. You will receive notification that your dataset has been submitted.
- A curator will review your dataset and if necessary, inform you about possible changes to be done before publication. The dataset might be returned to you to make these changes. Once the curator has approved your submission, you will be notified, and the dataset will be published and searchable as open access.
- If needed, you can modify metadata or data files after publication. Note that previous versions are not deleted, but archived as open access.
- **Please note:** Once you publish a dataset, you can't delete it. However, in certain cases, the dataset can be “deaccessioned,” which means that the data files no longer are openly available. But information about the dataset (author, title, description etc.) is still visible.

Retrieving your data

Once your dataset has been deposited and published, it can be cited as a reference using the information in the blue box on your dataset page. Your dataset is also assigned a Digital Object Identifier (DOI). This enables it to be easily found and increase its traceability.

Scholars Portal Dataverse > uOttawa - Dataverse > Making research data public: workshopping data curation for digital humanities projects >

Data Flow and Discovery Model

The screenshot shows a dataset page for "Data Flow and Discovery Model". At the top left, it says "Version 1.0". The main content area is a blue box containing the following text: "Tayler, Felicity; Mitchell, Marjorie, 2021, 'Data Flow and Discovery Model' Scholars Portal Dataverse, V1". To the right of this text is a red-bordered box containing the DOI: "https://doi.org/10.5683/SP2/VAPPJQ". Below the blue box, there is a "Cite Dataset" dropdown menu with options for "EndNote XML", "RIS", and "BibTeX". To the right of the blue box, there is a "Con" button and a "Dataset" label with "65 Dow" below it.

You can also grant editors or research partners access to your unpublished dataset.

References

The guidelines above have been adapted from several sources including:

University of Victoria Libraries. (2020, May 02). "Deposit Guidelines for UVic Dataverse." https://libguides.uvic.ca/ld.php?content_id=35154390

DataverseNO. "Deposit Guidelines." <https://site.uit.no/dataverseno/deposit/>

Research & Scholarship. (2020). How to deposit research data in the Agri-environmental Research Data Repository or the University of Guelph Research Data Repository (Research & Scholarship, Ed.; V4 ed.). Scholars Portal Dataverse. <https://doi.org/10.5683/SP2/CPHFGA>