McMaster Dataverse Data Deposit Guidelines

This document outlines requirements for researchers looking to deposit data in the McMaster Dataverse repository. McMaster Dataverse is part of Borealis, the Canadian Dataverse Repository.

A data curator at McMaster will review all deposited datasets for alignment with the Deposit Guidelines. Upon review the Depositor may be asked to make modifications to the dataset necessary to meet these guidelines. Curators will not request changes to the data itself, only metadata, file formats/organization, and data documentation.

Once a data curator has approved your submission, you will be notified, and the dataset will be published. All published datasets will receive a Digital Object Identifier (DOI) to allow your dataset to be cited.

A full walkthrough on how to deposit data to McMaster Dataverse can be found here.

Borealis maintains an advanced user guide for Dataverse. This can be helpful if researchers are interested in activating a guestbook, managing roles and permissions, working with APIs, setting up an embargo, and more specialized features.

Expectations for Depositors:

Many files may be created over the course of a research project. An important step for Depositors in preparing their data for deposit is the selection or appraisal of which files should be archived. In making these decisions, Depositors should provide enough supporting documentation for other researchers to understand how data were created, reproduce methods and findings, and reuse the data files.

When making decisions of what documentation to include with your data, consider what someone (or your future self) would need to know to understand, evaluate, analyze, or replicate your data without having to ask you.

Consult the resources available on the DataOne Data Management Skillbuilding Hub Best Practice: Describe website for more information on data documentation.

When possible, Depositors are strongly encouraged to include a version of their raw data, in addition to any processed data used in published analyses and figures. If raw data contain any sensitive information, Depositors should follow best practices for de-identification and ensure they have the proper permissions to share before depositing.

If your dataset is undergoing or will undergo a double-blind review, we can set up a fully anonymous dataset deposit. Please contact us directly at rdm@mcmaster.ca for details.

Standards for Deposit:

Before depositing in Dataverse, Depositors must make sure their dataset(s) meet the following standards.

1. Sensitive data must be <u>anonymized/de-identified.</u>

McMaster Dataverse does NOT accept content that contains confidential or sensitive information. Dataverse can be used to share de-identified and non-confidential data only. Contributors are required to remove, replace, or redact such information from datasets prior to upload.

If you have sensitive or confidential data that you need to store or share (containing personally identifiable information such as social security or credit card numbers, health records, etc.), please contact us directly to find out about the options for managing this kind of data outside of Dataverse or for help de-identifying and anonymizing your datasets.

If your data cannot be shared or is previously published, you can still upload metadata and code to create a persistent record of your research data and help others reproduce your findings.

Consult the following resource for help with de-identifying and anonymizing datasets.

- [De-identification Guidance document](#) from the Portage Network

2. Include a <u>ReadMe</u> file.

For research data to be read and interpreted correctly, it requires sufficient documentation. All deposited datasets must include a "ReadMe" file that includes the following information:

- Details about dataset creation
- Description of files contained in the dataset
- Information about dataset completeness
- Limitations on reuse

ReadMe files should be saved in a sustainable file format. This can include Unicode .txt files, rich text .rtf files, markdown .md file, or .pdf file.

Consult the following resource for a basic ReadMe file template with McMaster-specific guidance:

- [ReadMe template](#)

*<u>**NOTE**</u>: McMaster's template is a .docx file for clarity – please save this as a .txt or .rtf file for longevity when uploading it to Dataverse.*

3. Deposit your files in <u>sustainable file formats</u>.

In order to ensure the long-term accessibility of your data, use sustainable file formats. Sustainable file formats are standardized, open, well documented, in common usage, unencrypted, uncompressed, and lossless.

Some commonly used sustainable file formats include:

- Text: PDF/A, RTF, TXT, XML, LaTeX, md
- E-Books: EPUB, PDF/A
- Audio: FLAC
- Image: TIFF, SVG, JPEG 2000, BMP, PNG
- Medical Images: DICOM, NIfTI
- Spreadsheet: CSV, TAB
- Video: MP4, OGV, OGG, MJ2

Consult the following resources for a non-exhaustive list of preferred file formats.

- [Prepare your data: Preferred file formats](#) from DataverseNO.
- [Data Curation Primers](#) for various file formats from the Data Curation Networks.

4. Describe your dataset with rich <u>metadata</u>.

Depositors must complete all required fields in the citation metadata, including a link to a research publication where applicable. Depositors are strongly encouraged to complete geospatial metadata fields and subject-specific metadata fields, as appropriate. Consult the following resource for guidance on Dataverse metadata fields.

- Portage Network. Dataverse Metadata Best Practices Guide. https://portagenetwork.ca/wpcontent/uploads/2020/02/DVN_MetadataBestPracticesGuide_v2.0_ENG_20200219.pdf

Curators at McMaster Library may suggest changes to the descriptive metadata for the purposes of discovery, reuse, and preservation.

5. Use a consistent <u>file naming and organization scheme</u>.

A well-structured file hierarchy will make it easier to locate and share your files.

Following proper file naming conventions makes it easier to navigate and find specific files and allows other researchers to understand and reuse your dataset.

- Name files consistently
- Keep files names short (< 25 characters) but meaningful
- Do not use spaces to delimit words. Use capital letters, hyphens, or underscores
- Avoid special characters such as & , * % # * ( ) ! @$ ^ ~ ' { } [ ] ? < >
- Denote dates using ISO8601 standard YYYY-MM-DD (e.g. 2019-01-10).

Our commitment to your data:

McMaster University Library commits to preserving published datasets for a period of at least ten years from the date of publication. Our objective is the continued access and preservation of deposited datasets for the longer term. To support this objective, McMaster University Library reserves the right to convert deposited files to any medium or format and make multiple copies for the purposes of security, back up, and preservation. We will never modify file contents and only make changes to file formats in the interest of long-term access and reuse. Please note, we do not attempt to judge the scholarly quality of deposited datasets, and trust the judgement and research expertise of those who created and deposited the Dataset. Thus, a determination of a dataset's research quality is at the sole discretion of the Contact Person as named in descriptive metadata.

References:

The requirements and guidance above have been adapted from several sources, including:

- DataverseNO. DataverseNO Accession Policy. Version 2.1.
- DataverseNO. Deposit Guidelines: Prepare Your Data.
- Scholars Portal. Scholars Portal Dataverse Primer.
- Borealis. Borealis Dataverse Account Terms of Use.
- University of Victoria Libraries. Deposit Guidelines for UVic Dataverse