

## Guidelines for Recommending UBC Library Repositories for Data: cIRcle and Dataverse

---

Version 2.0. Created September 2016. Last updated February 2022

### About this Guide

UBC Library offers a number of services for managing and preserving research data. This guide was created and is maintained by members of the cIRcle and Research Data Management teams primarily to provide guidance to librarians around management of data deposit requests between cIRcle, UBC's open access digital repository, and UBC Dataverse Collection, a cross-disciplinary data repository supporting online management and access to data. To learn more about the major data repositories used by Canadian researchers including Dataverse, FRDR, and Dryad, see the [Where Should I Deposit My Data?](#) decision-tree. For more information about Research Data Management as a whole, please see [researchdata.library.ubc.ca](https://researchdata.library.ubc.ca)

### Overview

[UBC Library](#) defines research data as “[...] the data created or generated as part of a research project and exists in many formats including numeric data, text, transcripts, images, video and audio recordings.” (<https://casrai.org/rdm-glossary/>)

In recognition of the diversity of data and the complexity of data stewardship, UBC Library maintains or provides support for four repositories where research data may be stored, accessed, and managed: Dataverse, FRDR, Dryad, and [cIRcle](#) (DSpace). Each repository presents different strengths and limitations in the management of research data. The advantages of each are dictated, in large part, by the data creator's goals for access, preservation, and re-use. This document provides general guidance for librarians to decide primarily when to direct data deposits to cIRcle or the UBC Dataverse Collection. It also outlines how relationships between content living across both repositories may be identified and described.

### Guiding Principles

#### General

All four repositories can hold multiple files per record. Ideally, a record within a repository holds all appropriate files related to a study; where possible, multiple files from a study should not be split across multiple records within a repository. However, depending on the type of information deposited, a study may have a record for a publication in cIRcle, for example, and corresponding records for data in Dataverse, FRDR, or Dryad. This scenario is described in greater detail in the section: [When there is both Research Data and Documents](#)

Where materials for a study may require multiple records within a single repository or across different repository platforms, a consultation with any repository staff member will ensure best practices are applied.

## Final and Work-in-Progress Data

[cIRcle](#) is UBC's digital repository for teaching and research materials. cIRcle's mandate prioritizes open access and preservation for static or final versions of materials that will not require adjustments or changes such as articles, theses, and conference papers as well as recordings of events and lectures. Although cIRcle does accept datasets, its services offer limited support for version control. In most cases where a new version of a file becomes available, cIRcle generally recommends **adding** updated files to existing records rather than replacing (i.e. deleting) files. For this reason, cIRcle is an ideal repository for completed datasets, particularly when paired with an article, presentation or paper.

[Dataverse](#) is an open source web application for sharing, preserving, citing, exploring, and analyzing research data. It facilitates making data available to others, and allows replication of others' work. It has a robust version control and allows granular access to data files. Dataverse is an excellent place for files toward the end of the research data lifecycle or final version that might be amended by the users.

## Access & Management

While anyone can access materials deposited to cIRcle, [submissions](#) or edits to existing records are managed by [cIRcle staff](#) upon request. Although cIRcle does support embargoes, data should be added to cIRcle with the ultimate goal of open access and re-use. See the cIRcle [FAQ](#) for more information on permissions and withdrawal and editing policies.

Dataverse allows and encourages continuing use and further editing of research data, including version control and granular access to research data which requires a UBC CWL login.

## Data Type

Tabular data files can be deposited to either cIRcle or Dataverse depending on the data content and other considerations (final/in-progress, access, etc.).

SPSS (.por, .sav), Stata (.dta), R (.RData), LIDAR or GIS data would be more suited for deposit in Dataverse.

## File Sizes

cIRcle recommends individual file sizes not exceed **2GB**. For help with larger files, please [contact cIRcle staff](#).

In Dataverse there is a maximum file size of **2.5GB** for each file that users can upload via a browser. Larger files could be deposited by Dataverse staff via the API. For very large datasets (e.g. TBs), we recommend that data be deposited in FRDR. Contact the RDM team for help: [research.data@ubc.ca](mailto:research.data@ubc.ca).

## File Formats

Ideally, open and non-proprietary file formats should be used to store and share data. Please refer to [recommended file formats](#), and recommend conversion of file formats when possible.

ciRcle is file type agnostic, and can accommodate a [variety of file formats](#) though non-proprietary are preferred. Through [Open Collections](#), ciRcle is able to support in-browser viewing of documents (.pdf), images (.jpg, .png, .gif), videos (.mp4), and audio (.mp3).

Dataverse is file format agnostic, and can accommodate a variety of file formats. However, to make research data previewable and analyzable in a browser while using Dataverse, we recommend to upload/convert the data to CSV (.csv), SPSS (.por, .sav), Stata (.dta) or R (.RData) file formats.

Excel (.xls / .xlsx) may be referred to either repository.

## When there is both Research Data and Documents

Many researchers may wish to deposit both the research data for their study, as well as the publication or supplemental material. There are advantages and disadvantages to having all items for a study (regardless of format) within one repository.

ciRcle is an ideal home for publications with corresponding research data that is static and relatively small (under 2GB). It's also the place most users expect to find publications from UBC creators. Full-text indexing is provided through the multi-repository search portal, [Open Collections](#) for improved discoverability and access. Researchers may choose, however, to deposit their article in ciRcle and have links to their datasets in another data repository if file size, flexibility or interactivity requirements do not support having related materials in one place.

Dataverse is the preferred repository for researchers who want to deposit, describe and manage their datasets, particularly where versioning is a priority. For extremely large datasets (e.g. spectrometer data), FDRD is recommended.

## Linking between ciRcle and Dataverse

Related materials located in both ciRcle and Dataverse will have a note added to the record in each respective repository:

- In ciRcle, this information is currently included in the *Abstract/Description* field.
- In Dataverse, the *Related material* field is the best field to connect to ciRcle records.

See for example a study with a record in both ciRcle and Dataverse :

Honey-Rosés, J., Gill, D., & Pareja, C. (2016, March 3). *BC Municipal Water Survey 2016*.

- Publication in ciRcle: <http://hdl.handle.net/2429/57077>
- Data in Dataverse: <https://dx.doi.org/10.14288/1.0314341>

## Help

For questions about this document or research and data repository services, please contact us at

ciRcle at [circle.repository@ubc.ca](mailto:circle.repository@ubc.ca) or Research Data Management at [research.data@ubc.ca](mailto:research.data@ubc.ca).