

cIRcle-Dataverse Research Data Guidelines

Version 1.0, September 2016

Amber Saundry, Eugene Barsky and Paul Lesack

Overview

As described by [UBC Library](#), “Research data is 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.”

UBC Library provides access to two potential repositories for research data deposit: [cIRcle](#) (DSpace) and [Dataverse](#). Each repository provides different strengths and limitations in management of research data. This document is meant to provide guidance for how research data will be referred to and linked in each and across both repositories.

Guiding Principles

General

Both repositories can hold multiple files per record. Ideally, a record within a repository holds all appropriate files related to a study; multiple files from a study should not be split across multiple records within a repository. Depending on the type of information deposited, a study may have one record in Dataverse, and one record in cIRcle. This will be discussed further in the [When there is both Research Data and Documents](#) section.

There are cases where a study may recommend or require multiple records within a repository; please consult with appropriate Dataverse or cIRcle staff.

Final and Work-in-Progress Data

cIRcle’s main goals include access and preservation of material, and does not support versioning control. In general, cIRcle avoids replacement of files, but additional files can be added to a record at a later date. As a result, cIRcle is ideal for final datasets that will not require adjustment or undergo changes.

Dataverse is an open source web application to share, preserve, cite, explore and analyze research data. It facilitates making data available to others, and allows replicating others' work. It has a robust version control and allows granular access to data files. Dataverse is excellent place for work-in-progress data files or files toward the end of the research data lifecycle.

Access

cIRcle does not provide access to UBC researchers to edit existing submissions or replace existing files with newer versions. Researchers can however contact cIRcle staff to inquire about additions to an existing record on their behalf.

Dataverse allows and encourages continuing use of research data, including version control and granular access to research data. Dataverse also supports EZProxy for multiple school access points and soon could support Shibboleth for advanced access control (version 4.X)

File Sizes

For cIRcle it is recommended that the maximum file size not exceed 1GB. An overall record however, could have multiple files that total more than 1GB in size. If files are larger than 1GB, please [contact cIRcle staff](#).

In Dataverse there is a maximum file size of 2GB for each file that users can upload via a browser. Dataverse team can handle larger files via the backend of the system as needed.

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. 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 analyzable in a browser while using Dataverse, we recommend to upload/convert the data to SPSS (.por, .sav), Stata (.dta) or R (.RData) file formats.

Excel (.xls/xlsx) may be referred to either repository. If the data is chart-like in nature (cross-tabulations, etc.), then it may be more suited for cIRcle. If the data is similar to survey data (lines of data, each of which represent one entry), then it may be more suited for Dataverse.

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 ideal for the publication(s) corresponding with the research data. Full text indexing is provided through cIRcle, improving discoverability and use of material. It's also the place most users expect to find publications from UBC creators.

If the research data is in need of updating/version control or it is in a tabular format (which could allow this data to be analyzed in a browser), Dataverse is a great fit for this data. It may be recommended that the data be deposited in Dataverse, as opposed to within the cIRcle record holding the publication.

Linking between cIRcle-Dataverse Records

If it is deemed that it is advantageous for all items for a study to be split across cIRcle and Dataverse, links will be added to the record in each respective repository.

In cIRcle, this is currently included in the *Abstract/Description* field.

In Dataverse, the *Related material* field is the best field to connect to cIRcle records.

An example of a study with both a cIRcle and Dataverse record: 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: <http://hdl.handle.net/11272/10343>

Summary

- Tabular data files could be deposited to either repository, depending on the data content and other considerations (final/in-progress, access, etc.). SPSS (.por, .sav), Stata (.dta), R (.RData), LIDAR or GIS data may be more suited for deposit in Dataverse
- Final data files relating to publications should be deposited in cIRcle; work-in-progress data files should be deposited in Dataverse
- Linkage between cIRcle and Dataverse records should be implemented if data and publications are split across repositories
- Open Collections will expose all open data sets (cIRcle & Dataverse) via <https://open.library.ubc.ca/>
- When in doubt, please email UBC-CIRCLE@lists.ubc.ca and research.data@ubc.ca