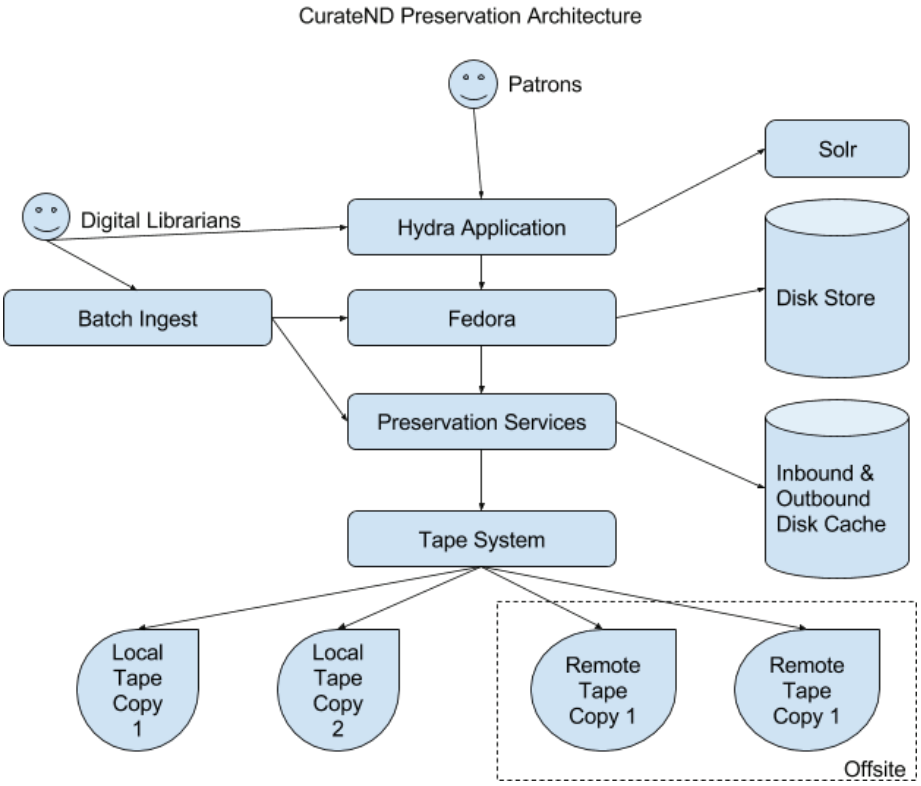
Data Curation Infrastructure

CurateND Data Curation Infrastructure

CurateND uses a Hydra-based discovery application. It uses Fedora Commons 3.x as the object registry and metadata store and Apache Solr as an index. Using both Fedora and Solr is common for Hydra applications. Self-deposit items go through the Hydra application. There is also a batch ingest ability, which deposits items directly into the preservation store as well as Fedora. Objects in Fedora contain pointers to our preservation store. The preservation store is a custom application that puts content into BagIt bags for storage on tape; maintains a disk cache of content; provides a URL for each preserved file; and runs fixity checks on the content. The data is ultimately all stored on tape, with two copies kept locally and two remotely. The tape appliance handles the replication.

Digital Librarians can deal with the batch ingest directly via a networked filesystem. Content is staged on the filesystem, where it can also be reviewed, assessed, and described. When it is ready, the librarian can start an ingest, which copies the data into the preservation system, the metadata into the preservation system, and a copy of the metadata into Fedora. It then asks the Hydra application to index the new content.


CurateND Preservation Architecture
Patrons
Solr
Digital Librarians
Hydra Application
Disk Store
Batch Ingest
Fedora
Preservation Services
Inbound &
Outbound
Disk Cache
Tape System
Local
Tape
Copy
1
Local
Tape
Copy
2
Remote
Tape
Copy 1
Remote
Tape
Copy 1
Offsite

Scholars Portal
A Service of the Ontario Council of University Libraries

Scholars Portal Home / Guides / Scholars Portal Dataverse Guide / About Dataverse

Scholars Portal Dataverse Guide

Search this Guide | Search

About Dataverse
- Scholars Portal Dataverse
- FAQ
- Related Resources
- Scholars Portal Dataverse Account Terms of Use

Your Dataverses

Your Research Data

Finding and Using Data

Help & Support

Go to Dataverse

Advanced Dataverse Guide

Dataverse Project Advanced User Guide

Library Contacts

Algoma University
Brock University
Carleton University
Lakehead University

Scholars Portal Dataverse

The Scholars Portal Dataverse is a repository primarily for research data collected by researchers and organizations affiliated with Ontario universities, although anyone in the world is welcome to use Scholars Portal Dataverse to deposit, share, and archive data.

Dataverse is an open-source tool developed by the Institute for Quantitative Social Sciences (IQSS) at Harvard University. Scholars Portal Dataverse is provided as a shared Ontario Council of University Libraries service.

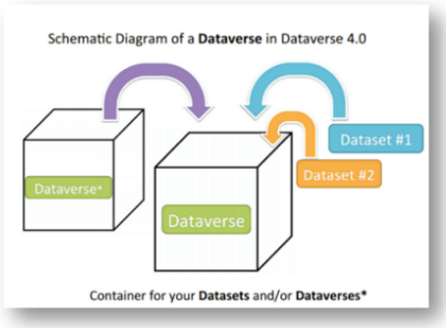
Researchers can use Dataverse to directly deposit data, create metadata, release and share data openly or privately, visualize and explore data, and search for data.

For more information about Scholars Portal Dataverse, please contact dataverse@scholarsportal.info

What is a Dataverse?

A Dataverse is a container for one or more Datasets or Dataverses. Each Ontario University has a Dataverse that contains many Dataverses and Datasets. Researchers can create Dataverses for their own research data and projects, and/or directly deposit Datasets within their Institutional Dataverse.

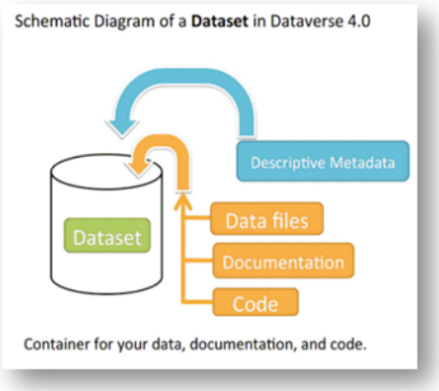
A Dataverse accepts all kinds of data files: tabular, text, image, etc. All file formats are accepted.



What is a Dataset?

A Dataset is a container for a particular research data set (this can include research data, code, and documentation).

Datasets have an associated metadata record (also referred to as cataloging information or data documentation). This metadata provides contextual information on the dataset. Please see here for more information on creating metadata for datasets.



Why use Dataverse?

Some key benefits to using Dataverse to manage your research data include:

- **Secure data management**. Dataverse supports the creation of terms of use and restrictions if you want to limit the use of or access to data. It also provides a backup copy for safekeeping.
- **Effective sharing**. Dataverse is a convenient way to disseminate your data, and can facilitate your research team's collaboration within a secure space.
- **Track changes.** Dataverse provides increased control over managing changes to a project without overwriting any part of that project, an especially useful feature when working on a team.
- **Long-term access and preservation**. Persistent identification to your data ensures reliable protection and prevention from data obsolescence.
- **Organization and compatibility**. Create your own personal web data archive that conforms to metadata standards to maximize system compatibility and searchability.
- **Save time**. Dataverse has an easy to use interface for uploading and searching through your data.
- **Increase research visibility**. Increase scholarly recognition for your work beyond your research publications.
- **Meet grant requirements**. Many funding agencies now *require* that researchers deposit data which collected as part of their research project into an archive.

References

Crosas M. The Dataverse Network: An Open-source Application for Sharing, Discovering and Preserving Data. D-Lib Magazine. 2011;Volume 17(1/2).

King, Gary. 2007. An Introduction to the Dataverse Network as an Infrastructure for Data Sharing. Sociological Methods and Research 36: 173–199. Available at http://j.mp/iHJcAa

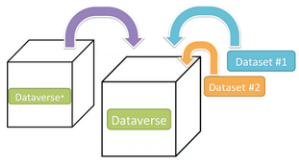
Texas Data Repository

About Documentation FAQs Log In Help

How Dataverse Works

A Dataverse is a container for datasets (research data, code, documentation, and metadata) and other Dataverses, which can be setup for individual researchers, departments, journals, and organizations.

Schematic Diagram of a **Dataverse** in Dataverse 4.0



Container for your **Datasets** and/or **Dataverses***

* Dataverses can now contain other Dataverses (this replaces Collections & Subnetworks)

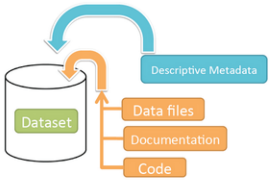
Each Dataverse contains datasets. Each dataset includes:

- Data files
- Metadata that describes the data files

Each dataset may also include:

- Code associated with the data files
- Additional documentation describing the data files and project form which they derive

Schematic Diagram of a **Dataset** in Dataverse 4.0



Container for your data, documentation, and code.

Help

- User Guide
- Helpdesk and Library Support

Hosted by

TDL.ORG
TEXAS DIGITAL LIBRARY