

# Introduction to Data Archiving with Borealis

At Scholars Portal using the Dataverse platform

# About Librarians / Archivists

- FYI – we can use your support
- We are seeking a fair deal from the employer
- Western needs more Librarians / Archivists
  
- Help us avoid a strike – potentially as early as **this Wednesday**
  
- **TAKE ACTION:** <https://www.uwofa.ca/take-action/>

# UWOF A

**Does your work as a faculty member or student include access to in-depth, subject-specific research and learning?**

If you answered **YES**, please **SUPPORT** your Librarians and Archivists whose work is vital for:

- ✓ ensuring access to cutting edge information technologies
- ✓ protecting the integrity of our libraries
- ✓ ensuring access to quality resources for teaching and research
- ✓ maintaining Western's top-ranked contribution to new scholarship

**...because librarians do it better than Google**

Librarians  
& Archivists

Western  LIBRARIES

# UWOFFA

**Librarians  
& Archivists**

**Western LIBRARIES**

## Bargaining 2023

We are still fighting for two important things: 1) complement (personnel numbers) and 2) a fair monetary package

- The Employer continues to reject any commitments to maintain let alone increase the complement of Librarians and Archivists. That means fewer people to help you.
- L&A Members deserve to be respected and valued. They are paid less at Western than other universities such as Brock, Waterloo, McMaster and Guelph
- The Bargaining Unit is small and the \$ we are for is asking for is well within what the Employer can afford. Workloads and student-to-Librarian-and-Archivist ratios are at an all time high, which actually saves the Employer money.
- Western is in a very strong financial position and came through the pandemic in a much better state than many other institutions.
- The monetary offer currently on the table is less than recent settlements in the sector. It does not make up for wage suppression or inflation.

# Background

# Tri-Agency Data Management Policy

- Policy includes 3 requirements:
  - Institutions: Institutional Strategy – “By March 1, 2023, research institutions subject to this requirement must post their RDM strategies and notify the agencies when they have been completed.”
  - Researchers: Data Management Plans – “By spring 2022, the agencies will identify the initial set of funding opportunities subject to the DMP requirement.”
  - Researchers: Data Deposit – “After reviewing the institutional strategies and in line with the readiness of the Canadian research community, the agencies will phase in the deposit requirement.”

# Data deposit

- In cases where journal publications, pre-prints and other research outputs arise from agency-supported research, researchers may be required to deposit research data and code in an appropriate repository
  - The code is a crucial addition – focus here is not so much on reuse of the data but reproducibility of the results
- Goals:
  - Ensure proper use of public funds
  - Facilitate reproducibility of results
  - Facilitate data sharing

# FAIR Principles for Open Data

- **Findable**

- The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets

- **Accessible**

- Once the user finds the required data, they need to know how they can be accessed, possibly including authentication and authorization.

- **Interoperable**

- The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- **Reusable**

- The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- <http://go-fair.org/>

# Concerns

- The concept of required data sharing is something that many researchers find problematic. Ethical, ownership and data sovereignty can all be legitimate barriers to data sharing, and the Tri-Agencies recognize this.
- Data deposit does not necessarily mean data sharing
- If you as a researcher think in good faith that there is a significant privacy, security or ownership issue that makes sharing inadvisable... **you are probably right**. Explain in your Data Management Plan and outline how you will handle preservation without sharing or explain why preservation is inadvisable. Funding agencies are aware that there are many cases in which data cannot or should not be shared.

# Dataverse and Borealis

# What is Dataverse? What is Borealis?

- Dataverse is a research data platform and repository for sharing, discovering and preserving research data
  - An open source project of Harvard University's Institute for Quantitative Social Sciences. Formal development began in 2006, based on earlier projects dating back to 1987
- Many international collaborators and used world-wide for all kinds of data (not just social science!)
- Borealis is a Canadian implementation of Dataverse currently hosted by the Ontario Council of University Libraries and supported by the Digital Research Alliance of Canada as an approved repository
- Dataverse is the name for the software, for the hosted archive and for individual collections of data files – Western University's Dataverse runs on the Dataverse installation at Scholar's Portal, and contains several researcher dataverses

## 63 Installations

Dataverse  
installations around  
the world



# Some benefits of Borealis for researchers

- Web-based tool with no installation required
- Meet grant & journal requirements for sharing research data, compliant with Tri-Agency requirements
- Share research data publicly, privately, within a single institution, or within a single research team
- Store data for long-term access
- Create a persistent identifier (DOI) for data
- Track versions of data over time
- Automatically generate citations for datasets

# The Borealis project at Scholar's Portal

- Scholars Portal is a collaborative project of the Ontario Council of University Libraries to provide shared infrastructure, collections and technology
- The SP Dataverse project began in 2012. Housed on University of Toronto servers to provide fully Canadian-hosted data archiving
- Local development includes French internationalization project with University of Montreal, integration with DataCite Canada, etc. Hosts dataverses for institutions across Canada
- Recently rebranded Borealis: a nationally coordinated, standardized data archiving service using Dataverse

# Key software features

- Supports uploading of any file type, including shapefiles, images, text, tabular data and code
- Certain types of tabular data files (Stata, SPSS, Excel, csv) are ingested into the data explorer and can be viewed online using the Data Explorer
- Geospatial shapefiles can be further explored and manipulated through integration with [WorldMap](#)
- Data can be made available openly, shared by request only, or only shared within a specific research team or user group
- Securely backed up, versioning control allows you to access old versions
- Generates a standard data citation (DOI) for use in research publications
- Author retains control and can choose from several default licenses

# Key limitations

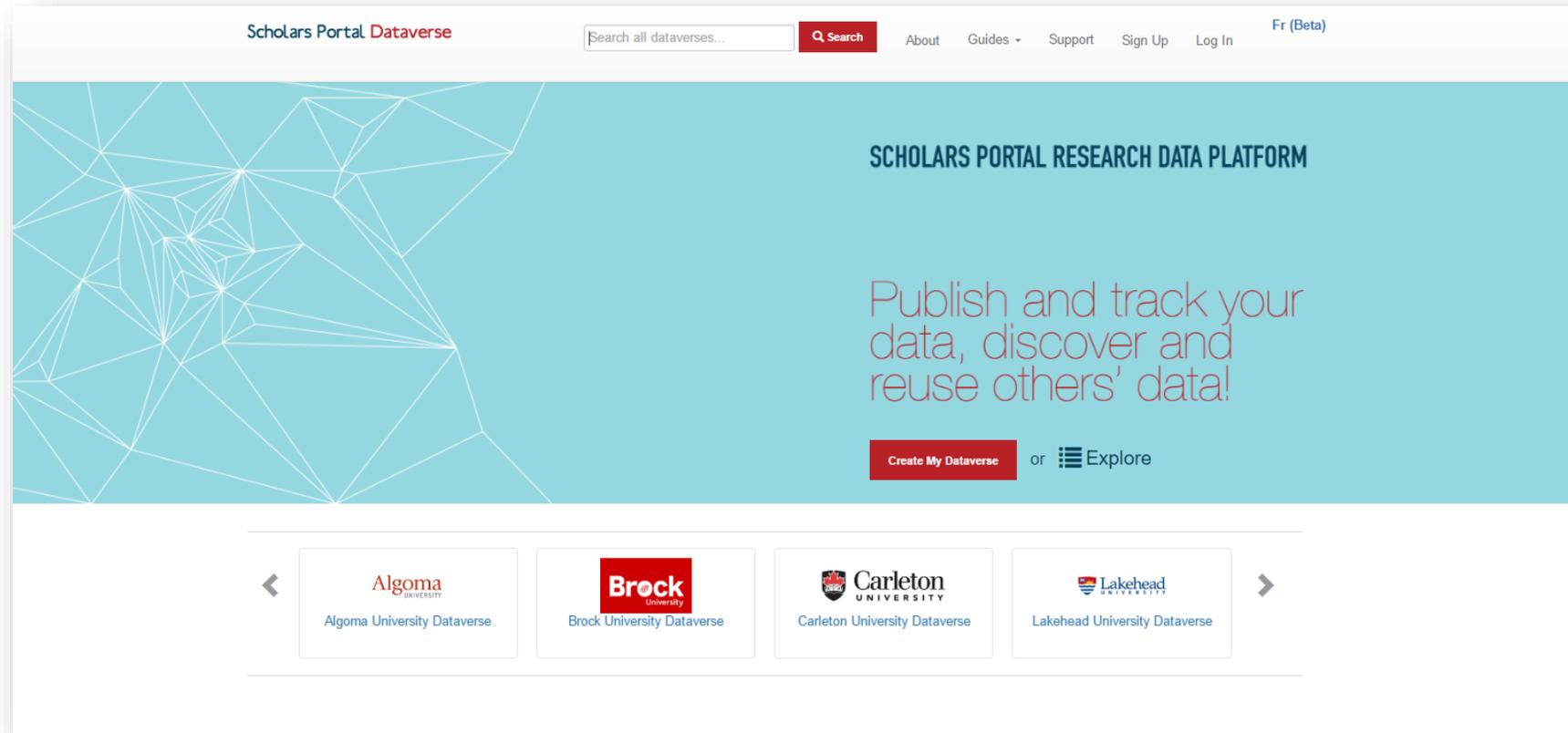
- Current file size limit is 3.5 GB (although changes regularly and exceptions are possible). For larger datasets (into the terabytes) consider Canada's [Federated Research Data Repository \(FRDR\)](#)
- Not suitable for confidential data – while you can restrict access to files, it is not approved for data with ethical concerns around sharing and they **really** do **not** want your confidential data
- File ingest doesn't deal nicely with Excel data that doesn't fit its expectations for a tabular file
- While data can be updated, removing a dataset or Dataverse is strongly discouraged and can only be done by platform administrators. Make sure you know what you are doing before you hit “publish”

# Publishing control

- Choose your access settings: modify access to your data files - ranging from fully open to completely restricted.
- Customize the terms of use, specifying how your data can be used, using one of our default licenses or a custom one.
- Modify access to your data files at any time.
- Once your dataset has been published, your metadata is publicly available. Therefore, your data is always discoverable.
  - If you don't want your data to be discoverable but need to share your data (e.g. within your research group or to meet a journal requirement) keep the file unpublished. Data will be available at a private url but not discoverable.

# Sharing your data

# Getting Started



# Log in

Scholars Portal **Dataverse** Search User Guide Support English Log In

## Log In

Log in or sign up with your institutional account — more [information about account creation](#). Leaving your institution? Please contact [Dataverse Support](#) for assistance.

**Your Institution** University of Western Ontario

Please select... Continue

Institution not listed

Username/Email

- Please select...
- Carleton University
- Concordia University
- McGill University
- McMaster University
- OCAD University
- Polytechnique Montréal
- Queen's University
- Université du Québec à Chicoutimi
- University of British Columbia
- University of Guelph
- University of Regina
- University of Toronto
- University of Victoria
- University of Western Ontario**
- University of Windsor
- UQAM | Université du Québec à Montréal
- Wilfrid Laurier University

Copyright © 2020

Powered by **The Dataverse Project** v. v4.19.3-SP

# Using your Western credentials

Western WebLogin

Log In [About WebLogin](#) [FAQ](#)

User ID:

Password:

[Log In](#)

# Western data repository

- <https://borealisdata.ca/dataverse/westernu>

 **Dataverses (14)**

 **Datasets (30)**

 **Files (730)**

**Dataverse Category**

Research Project (4)

Research Group (3)

Researcher (2)

Organization or Institution (1)

---

**Publication Year**

2021 (6)

2020 (9)

2019 (9)

2018 (3)

2015 (5)

[More...](#)

---

**Publication Status**

Published (37)

Unpublished (6)

Draft (4)

In Review (1)

---

**Author Name**

Taylor, Zack (4)

Dunn, Linda (3)

Sharawy, Mahmoud (3)

Heit, Bryan (2)

Johnson, Erin (2)

1 to 10 of 44 Results Sort ▾

**Dataset for common bean (*Phaseolus vulgaris* L.) with increased cysteine and methionine concentration** Draft Unpublished 

May 17, 2021 - Frédéric Marsolais Dataverse

 Viscarra-Torrico, Rosa Cecilia; Pajak, Aga; Soler Garzón, Alvaro; Zhang, BaiLing; Pandurangan, Sudhakar; Marwan Diapari; Song, Qijian; Miklas, Phillip; Hou, Anfu; Marsolais, Frédéric, 2021, "Dataset for common bean (*Phaseolus vulgaris* L.) with increased cysteine and methionine concentration", <https://doi.org/10.5683/SP2/JL7Y2I>, Scholars Portal Dataverse, DRAFT VERSION, UNF:6:9S9YcoqN9q4LFT6L9Jk38Q== [fileUNF]

Dataset1\_SupplInfo. SDS-PAGE of protein profiles of parents and recombinant inbred lines. M: Morden003; S: SMARC1N-PN1. Pop1 is indicated by the prefix F08-18-01 followed by a number; Pop2 by the prefix F08-18-02; and Pop3 by the prefix W09-01-01. The protein genotype is indicated...

---

**Frédéric Marsolais Dataverse** (University of Western Ontario) Unpublished 

 Apr 19, 2021

---

**Cognitive testing and questionnaire data from survivors of COVID-19.** 

Apr 16, 2021 - OwenLab Dataverse @ Western

 Wild, Conor J.; Norton, Loretta; Swartz, Richard H; Owen, Adrian M., 2021, "Cognitive testing and questionnaire data from survivors of COVID-19.", <https://doi.org/10.5683/SP2/ZQR9QQ>, Scholars Portal Dataverse, V1, UNF:6:kA9D/3RKDxJ3ua0CdiFT3Q== [fileUNF]

These data comprise the cognitive test and survey data used in the publication, "Seeing through brain fog: disentangling the cognitive, physical, and mental health sequelae of COVID-19." In this online study, we recruited participants from around the world who reported having had...

---

**OwenLab Dataverse @ Western** (University of Western Ontario) 

 Apr 16, 2021

This dataverse contains data, files, scripts, etc. from researchers in the OwenLab (<https://owenlab.uwo.ca/>) at Western University.

---

**Barnett Shakespeare Book Inventory** Draft Unpublished 

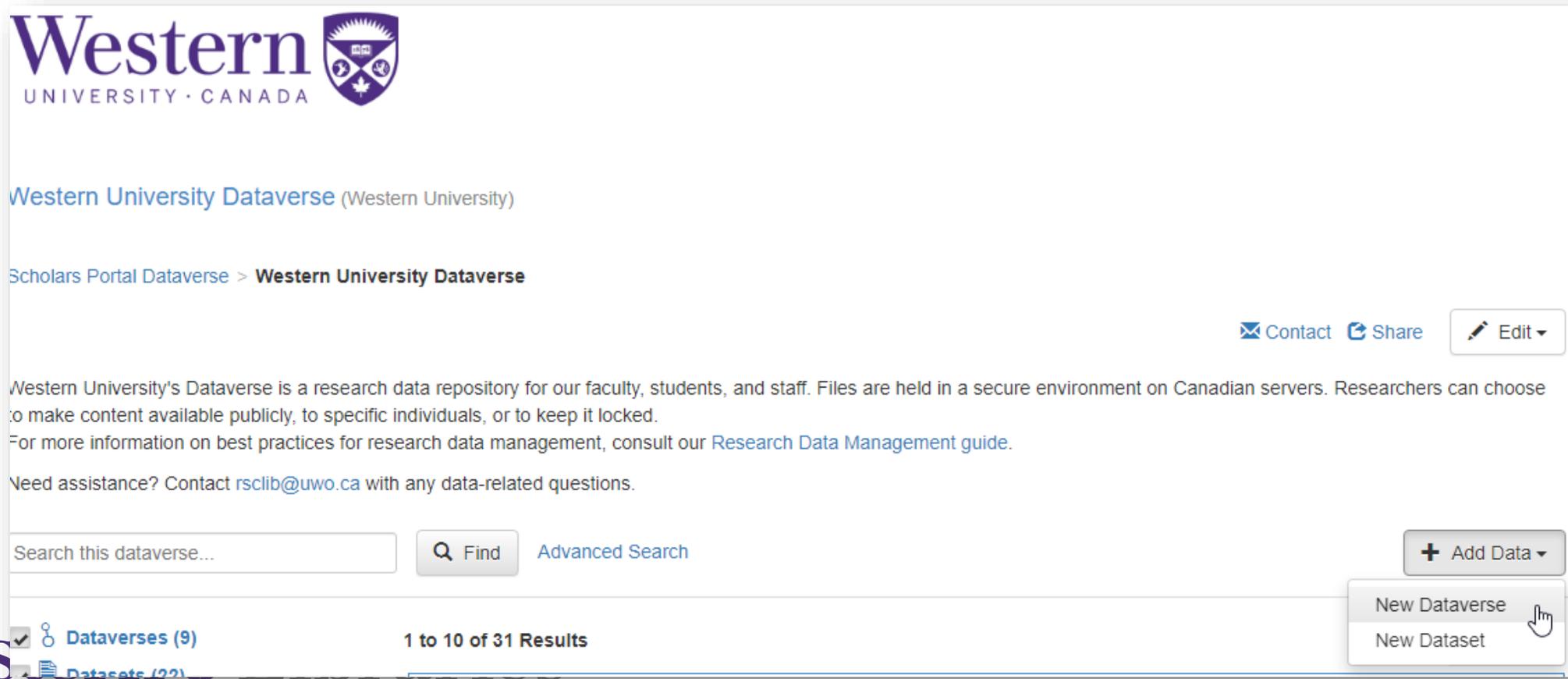
 Feb 3, 2021 - Barnett Shakespeare Project

# In-line file types

- Allows users to view certain file types directly in the browser without needing to download the file
- File types include:
  - Text and document files: txt, html, pdf
  - Audio files: mp3, mpeg, wav
  - Image files: gif, jpeg, png
  - Video files: mp4, quicktime
  - Statistical formats: Stata, R

# Adding a new data collection

- Create a new Dataverse to hold your data – don't just try to dump it in the root Western Dataverse



The screenshot displays the Western University Dataverse homepage. At the top left is the Western University logo. Below it, the text reads "Western University Dataverse (Western University)". A breadcrumb trail shows "Scholars Portal Dataverse > Western University Dataverse". On the right side, there are buttons for "Contact", "Share", and "Edit". The main content area contains a paragraph describing the Dataverse as a research data repository and a link to the "Research Data Management guide". Below this is a search bar with the text "Search this dataverse...", a "Find" button, and a link to "Advanced Search". At the bottom left, there are filters for "Dataverses (9)" and "Datasets (22)". In the bottom right corner, a "+ Add Data" button is open, showing a dropdown menu with two options: "New Dataverse" and "New Dataset". A mouse cursor is pointing at the "New Dataverse" option.

# Creating your Dataverse

- Basic metadata

Kristi Thompson Dataverse (University of Western Ontario) Unpublished

Scholars Portal Dataverse > Western University Dataverse > **New Dataverse**

\*Asterisks indicate required fields

**Host Dataverse** ⓘ

Western University Dataverse

**Dataverse Name** \* ⓘ

Kristi Thompson Dataverse

**Identifier** \* ⓘ

https://dataverse.scholarsportal.info/dataverse/ kthompson

**Category** \* ⓘ

Select one... ▼

- Select one...
- Department
- Journal
- Laboratory
- Organization or Institution
- Researcher**
- Research Group
- Research Project
- Teaching Course
- Uncategorized

**Affiliation** ⓘ

University of Western Ontario

**Description** ⓘ

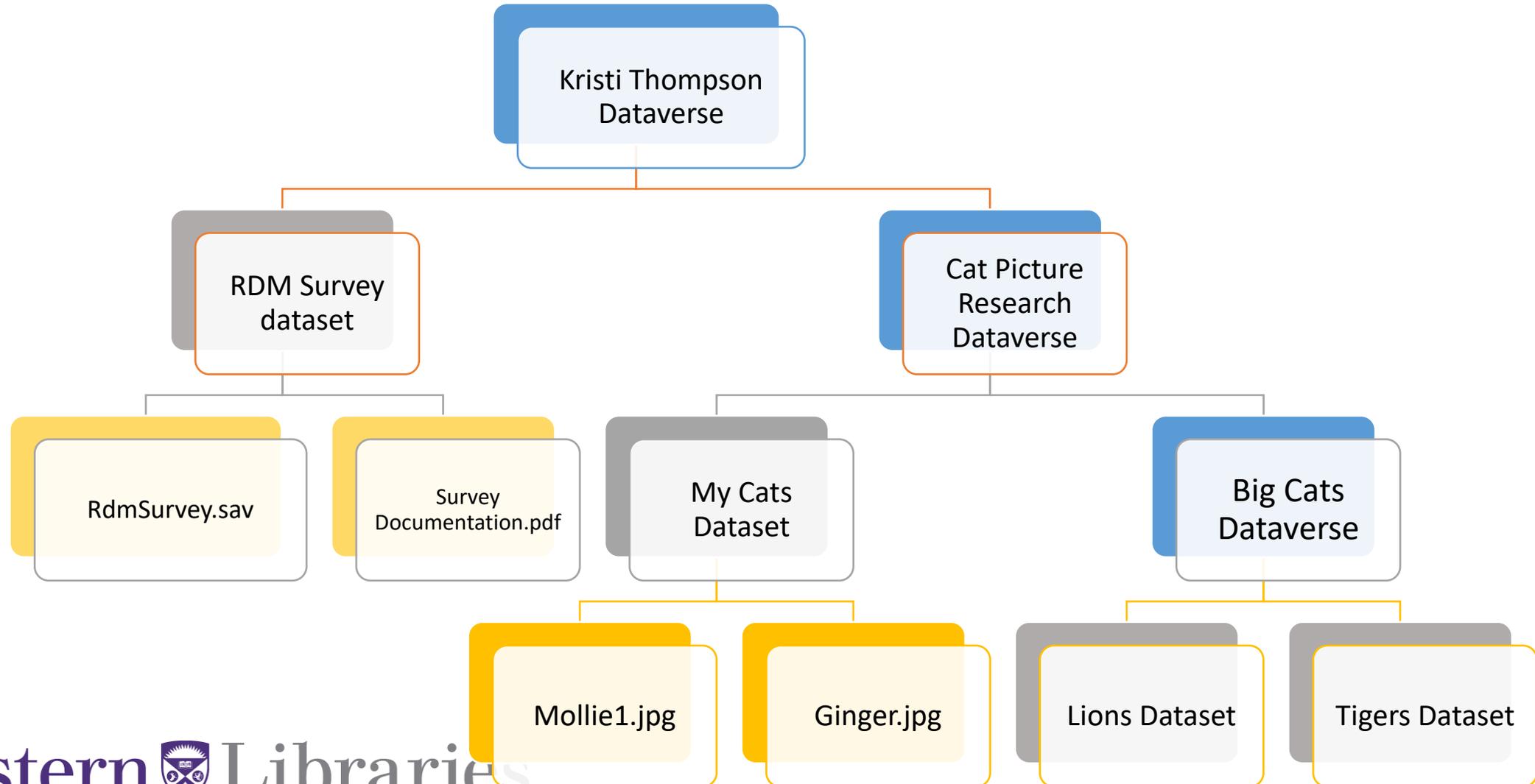
This field supports only certain HTML tags.

Choose the metadata fields to use in dataset templates and when adding a dataset to this dataverse

# Organizing with sub-Dataverses

- I might decide to keep all my datasets in one “Kristi Thompson” researcher Dataverse
- Each dataset you create can have multiple files associated with it
  - In theory you could create one dataset as a holder for a bunch of unrelated files, but that would be weird...
- If you have multiple unrelated research projects you might create a Dataverse for each of them
- If you have a very complicated research life you could even create several levels of Dataverse, treating it like a folder hierarchy
- Whatever you do, make up your mind before publishing. After a dataset or Dataverse has been published, it can only be moved by an employee of our host at Scholars Portal

# Researcher Dataverse organization



# Adding data

Scholars Portal Dataverse > Western University Dataverse > **Kristi Thompson Dataverse**

 **Success!** – You have successfully created your dataverse! To learn more about what you can do with your dataverse, check out the [User Guide](#).

[Contact](#) [Share](#) [Publish](#) [Edit](#)

Search this dataverse... [Find](#) [Advanced Search](#) [+ Add Data](#)

[Dataverses \(0\)](#)  
 [Datasets \(0\)](#)  
 [Files \(0\)](#)

This dataverse currently has no dataverses, datasets, or files. You can add to it by using the Add Data button

- New Dataverse
- New Dataset

# Data ingest

- If you upload [tabular data](#) (e.g., SPSS, STATA, RData, CSV, and Excel), the Dataverse software will ingest your tabular files and convert them to a tab-delimited file
- Ingested tabular data can be exported to / downloaded as a variety of formats including Excel, SPSS, Stata
- The pre-ingested tabular file you uploaded is saved within the repository and is available for download. However, the download method may not be obvious to users so consider adding a note if you want people to do this.

# Ingest Oddities

- ZIP (compressed) files can be uploaded to your dataset using the same method as any other file format. However, when the upload is complete, each file within the ZIP file will be listed, instead of a single ZIP file. The Dataverse software unzips the first layer of compression during the upload process (if you have a zipped zip file, it will stay zipped)
- The ingest process will only process the first tab in a spreadsheet file. All other tabs will be ignored and will not be included in the tabular file that's created. If you want all the tabs in a spreadsheet file to be converted to the TAB format, you'll need to upload each tab as its own file.

# Sharing unpublished data

## Barnett Shakespeare Project

(University of Western Ontario) **Unpublished**

Borealis > Western University Dataverse > Barnett Shakespeare Project >

**!** Info – This draft version needs to be published. When ready for sharing, please **publish** it so that others can see these changes.

**i** Unpublished Dataset Private URL – Privately share this dataset before it is published: <https://borealisdata.ca/privateurl.xhtml?token=>

## Barnett Shakespeare Book Inventory

**Draft** **Unpublished**



James, Blake, 2021, "Barnett Shakespeare Book Inventory", <https://doi.org/10.5683/SP2/PF7ORY>, Borealis, DRAFT VERSION, UNF:6:pmOALRSUqo0WoDNFHbmjQ== [fileUNF] **?**

Cite Dataset **▼**

[Learn about Data Citation Standards.](#)

Access Dataset **▼**

Publish Dataset **▼**

Edit Dataset **▼**

Contact Owner

Share

# Metadata!

Citation Metadata ^

**Title** \* ?

**Author** \* ?

**Name** \* ?

**Affiliation** ?  +

**Identifier Scheme** ?

**Identifier** ?

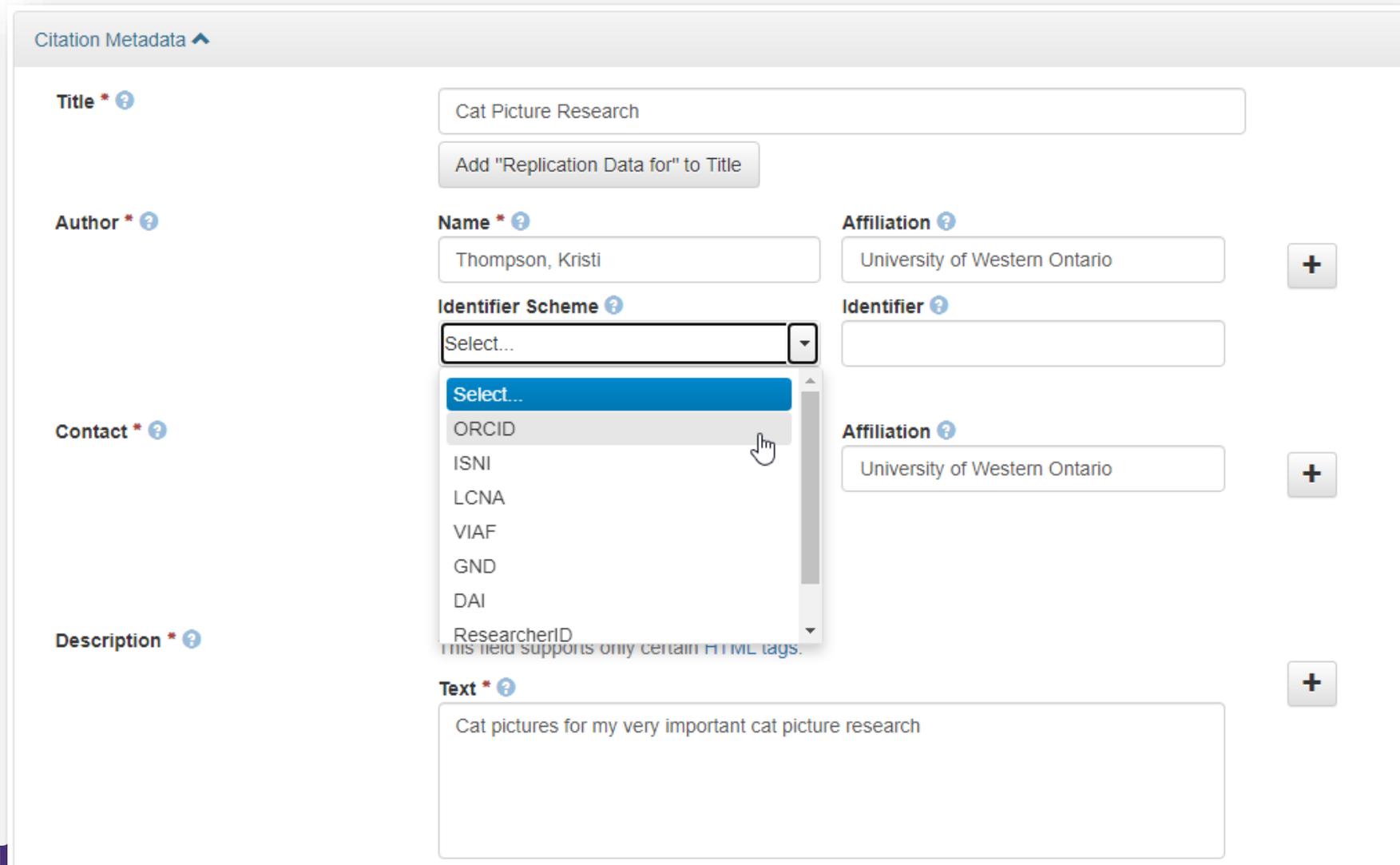
**Contact** \* ?

**Affiliation** ?  +

**Description** \* ?

**Text** \* ?

This field supports only certain HTML tags.

A screenshot of a web-based citation metadata form. The form is titled "Citation Metadata" and contains several fields for entering information. The "Title" field contains "Cat Picture Research" and has a button to add "Replication Data for" to the title. The "Author" section includes a "Name" field with "Thompson, Kristi" and an "Affiliation" field with "University of Western Ontario". The "Identifier" section has a dropdown menu for "Identifier Scheme" which is currently open, showing options like "ORCID", "ISNI", "LCNA", "VIAF", "GND", "DAI", and "ResearcherID". The "Identifier" field is empty. The "Contact" section has an "Affiliation" field with "University of Western Ontario". The "Description" section has a "Text" field with "Cat pictures for my very important cat picture research". There are plus signs next to the affiliation fields, suggesting they can be added or removed. A small note at the bottom of the form states "This field supports only certain HTML tags."

# More metadata!

Subject \* ?

Keyword ?

Related Publication ?

Notes ?

ID Type ?

ID Number ?

URL ?

Enter full URL, starting with http://

Agricultural Sciences

Arts and Humanities

Astronomy and Astrophysics

Business and Management

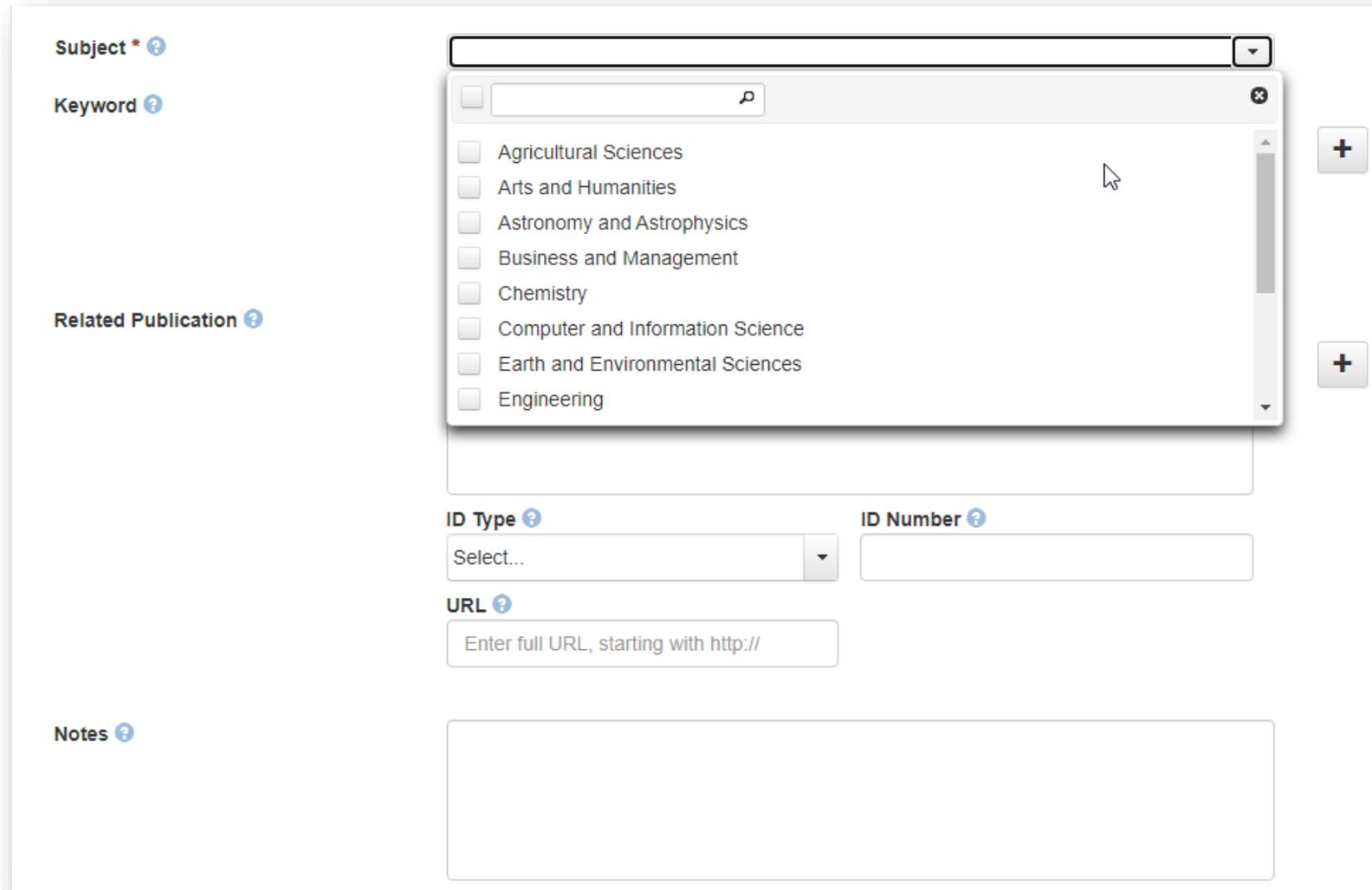
Chemistry

Computer and Information Science

Earth and Environmental Sciences

Engineering

+ +

A screenshot of a metadata form interface. The form is divided into several sections: 'Subject \* ?' with a dropdown menu; 'Keyword ?' with a search input; 'Related Publication ?' with a search input; 'Notes ?' with a text area; 'ID Type ?' with a dropdown menu; 'ID Number ?' with a text input; and 'URL ?' with a text input containing the placeholder 'Enter full URL, starting with http://'. A subject dropdown menu is open, showing a list of categories with checkboxes: Agricultural Sciences, Arts and Humanities, Astronomy and Astrophysics, Business and Management, Chemistry, Computer and Information Science, Earth and Environmental Sciences, and Engineering. To the right of the list are two '+' buttons. The form is presented with a slight drop shadow.

Optional metadata – only accessible after you save your dataset and then go back to “edit”

Characteristic of Sources Noted [?](#)

Documentation and Access to Sources [?](#)

Geospatial Metadata [▼](#)

Social Science and Humanities Metadata [▼](#)

Astronomy and Astrophysics Metadata [▼](#)

Life Sciences Metadata [▼](#)

Journal Metadata [▼](#)

# Do you really love metadata?

- [Canadian Dataverse Metadata Best Practices Guide](#)

Citation Metadata

Citation Metadata Block

Field	Definition with tips	Required/ Recommended/ Optional	Example
<b>Title</b>	Full title by which the Dataset is known.	required	Social Media Use Among Teens, 2015 [Canada]
<b>Subtitle</b>	A secondary title used to amplify or state certain limitations on the main title. <i>Tip: subtitle is not included in generated citation. Include subtitle with title to be included in citation.</i>	recommended (if applicable)	Main Survey
<b>Alternative Title</b>	A title by which the work is commonly referred or an abbreviation of the title. <i>Tip: Acronym, short form or translation of full title.</i>	optional	Youth Social Media Survey
<b>Alternative URL</b>	A URL where the Dataset can be viewed, such as a personal or project website.	optional	<a href="http://youthsocialmedia.org">http://youthsocialmedia.org</a>
<b>Other ID</b>	Another unique identifier that identifies this Dataset (e.g., producer's or another repository's number). Consists of 2 subfields.		
Agency	Name of agency that generated this identifier.	optional	Youth Communication Development Project, Ec Department, Queen's University

# Applying a license

- Default to CC0 (open data) OR:
  - Templates for Create Commons Licences
  - Custom terms of use 'Data Usage Agreement'
- Control access to your data at the dataset and file level:
  - Assign permissions for collaborators, curators, file downloaders (access)
  - Granular file-level permissions
  - Restricted files and custom terms of access

# Publishing your Dataverse and datasets

- After creating your Dataverse you can customize the appearance with header images, custom text and links, and create a guestbook
- A new Western Dataverse user can create unpublished Dataverses and datasets within Western's Dataverse and submit them for publication, but the administrator of the Western Dataverse (that's me!) needs to approve them before they are made public
- Remember – once it's published, you can't delete it!



# Why default to CCO (and why can't I delete it?)

- Dataverse supports the ideal of “FAIR” data
- **Findable**
  - Open archive, documented, searchable
- **Accessible**
  - Persistent identifier, always remains available – at least metadata even if data is withdrawn
- **Interoperable**
  - Machine readable, standard formats
- **Reusable**
  - Well-documented, understandable, license allows for re-use

FAIR Data:

“As open as possible, as closed as necessary.”

## Links and resources:

- [Western University's Data Archive on Borealis](#)
- [Western Research Data Management Guide](#)
- [Borealis User Guide](#)
- [Dataverse North Metadata Best Practices Guide](#)

For more information or to book an appointment to discuss your data, contact Western Libraries Research and Scholarly Communications Team, [rsclib@uwo.ca](mailto:rsclib@uwo.ca)

- Research Data Librarian Kristi Thompson
- Data Librarian Liz Hill