

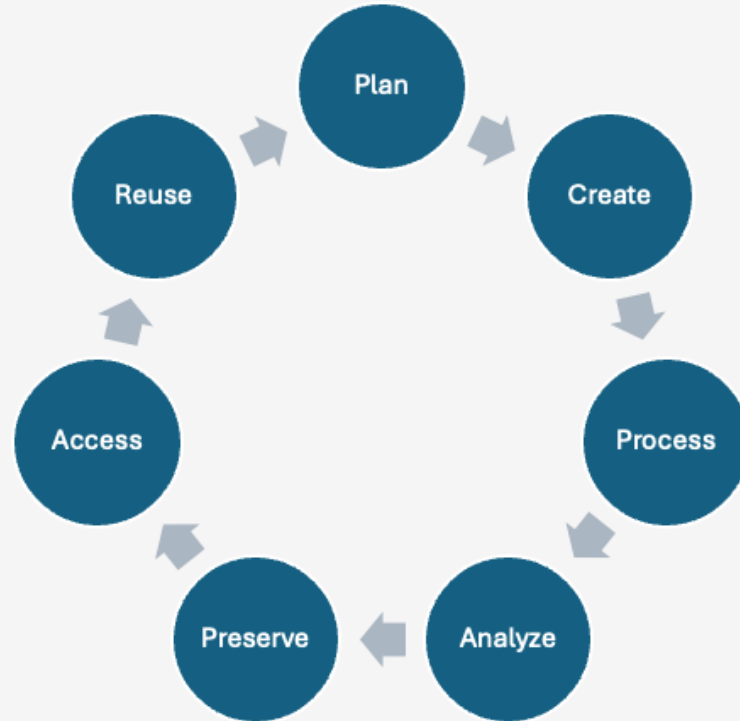


[source](#)

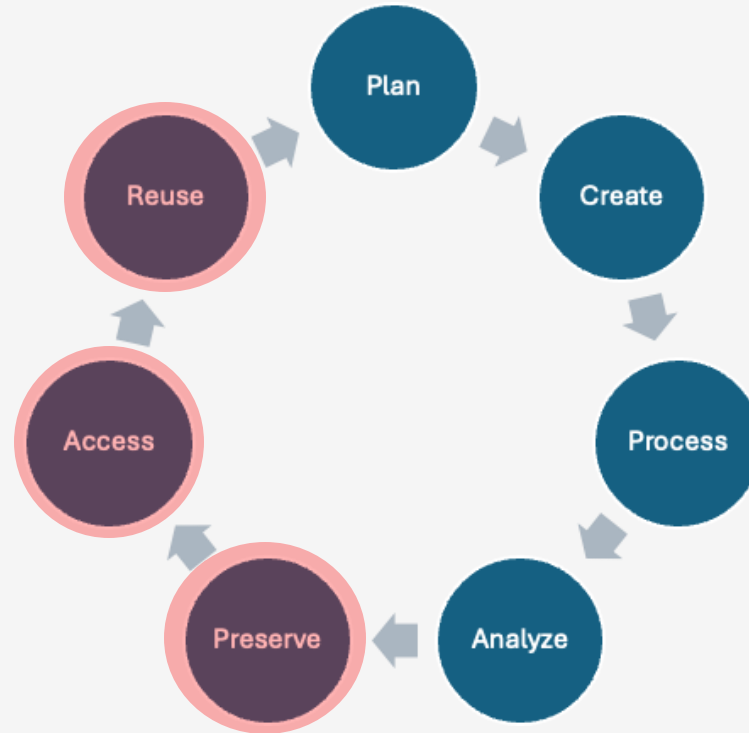
Data Deposit and Sharing

Nick Rochlin, Data Science Librarian
UVic Libraries

Research Data Lifecycle



Research Data Lifecycle





SHARING
It's not for everyone

Sharing Data

Why would you share data?

Why would you share data?

- Because someone told you to (publishers, funders, supervisors, etc.)

Why would you share data?

- Because someone told you to (publishers, funders, supervisors, etc.)
- Because it's good research practice
 - Enables research to be reproduced
 - Allows others to do new work based on what you've done (with proper attribution)
 - Allows others to check for errors or bias
 - Supports the spirit of the FAIR principles

What to Share vs. What to Keep



What should you share?

- ****Data**
 - Raw data (directly obtained from instruction, simulation, survey, etc.)
 - Processed data (cleaned or manipulated as a step prior to analysis)
 - Analyzed or summary data (results of analysis of the processed data)
 - Other?

What should you share?

- ****Data**
 - Raw data (directly obtained from instruction, simulation, survey, etc.)
 - Processed data (cleaned or manipulated as a step prior to analysis)
 - Analyzed or summary data (results of analysis of the processed data)
 - Other?
- Documentation that supports reuse of data
 - README
 - Data dictionary
 - Codebook
- Code
 - Cleaning
 - Analysis
 - Visualization

What should you share?

- Anything else?

What to Keep for Yourself?



What to Keep for Yourself?

- Project management records: internal communications, decision making documents, admin records, ethics information, etc.
- Unused analyses, visualizations, code chunks...
- Anything else?

Disclaimer

Just because you're not sharing something doesn't mean you should get rid of it

- Most institutions have rules for how long you have to keep research records after a project is complete
- REBs don't always require you to destroy data, but if you tell participants their data will be destroyed, then you must do it
- Consider all these things prior to starting a project (DMP plug!)

Disclaimer Part 2

Make sure that everybody involved in the projects agrees on what data are being shared (if any), and how/where that will happen

- This can include:
 - Your supervisor
 - Partners/collaborators/communities you're working with
 - Research participants

How to Share Data

- On request
- Via a website that you maintain
- **Via a data repository**

Reasons Data Isn't Shared



Reasons Data Isn't Shared

- Sensitivities
- Intellectual property
- Community data
- Too much data
- Time, effort, and priorities



Data Repositories

What is a Data Repository?

- Systems that facilitate the upload, long-term preservation, discovery, and access of research data and accompanying materials
- Focus on housing data **after a project's completion**, and that won't be changing on a regular basis

Why Use a Data Repository?

- You don't have to handle all aspects of data sharing yourself, eg:
 - Responding to data requests
 - Maintaining a website
 - Data integrity checks over long periods of time
 - ...
- You can get help from repository support staff
- You can get metrics about how often your dataset is viewed/downloaded

Things to Look for in a Data Repository

- Good data repositories will:
 - Have a mandate and plan for data preservation
 - Offer information about the data set that enables people to discover and learn about the data
 - Provide either direct access to the data or information on access conditions
 - Ensure that datasets have a persistent identifier that will provide perpetual access to the dataset (no broken links)



Types of Data Repositories

- **Generalist:** Accept data from any discipline
- **Domain-specific:** Limited to data from specific fields or disciplines (e.g. genomics, social sciences, etc.)
- **Institutional:** Only accept data from people affiliated with a specific institution

Borealis

- Shared service with Canadian regional library consortia, academic institutions, research organizations, and the Digital Research Alliance of Canada
- Built on open-source software, The Dataverse Project (Harvard), hosted by Scholars Portal at the U of T Libraries
- Data stored on the Ontario Library Research Cloud

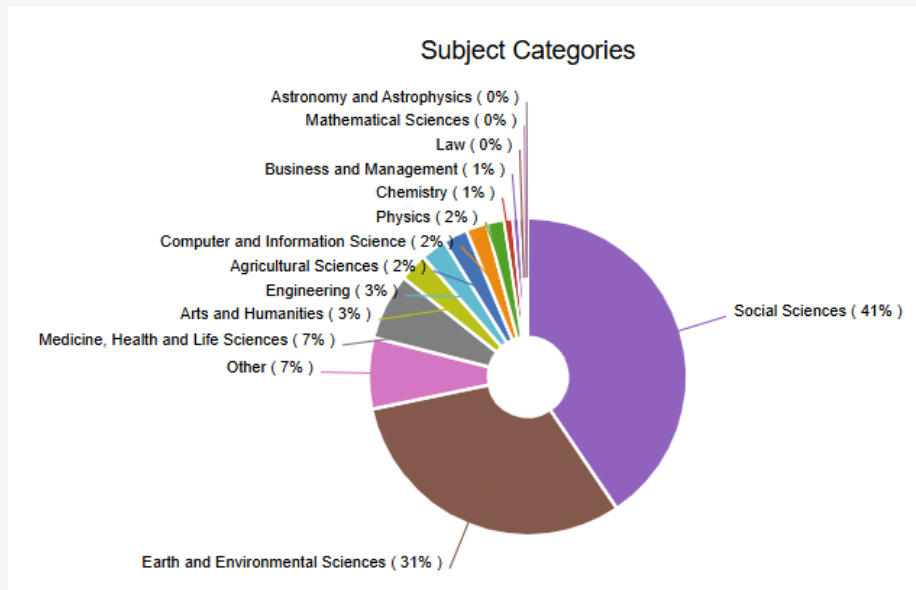


borealis

The Canadian Dataverse Repository
Le dépôt Dataverse canadien

Borealis

- 70+ member institutions
- Just under 23,000 datasets
- Over 25M downloads



UVic Borealis

- Deposits began in 2016
 - 294 datasets
 - 28 Dataverses
 - 5,842 files

Deposit Date

2025 (38)
2024 (123)
2023 (20)
2022 (21)
2021 (31)
2020 (24)
2019 (11)
2018 (9)
2017 (7)
2016 (4)

Subject

Other (148)
Earth and Environmental Sciences (125)
Medicine, Health and Life Sciences (35)
Social Sciences (21)
Arts and Humanities (18)
Computer and Information Science (11)
Physics (10)
Chemistry (6)
Engineering (5)
Law (4)
Mathematical Sciences (4)
Business and Management (3)
Agricultural Sciences (2)

Semi-mediated curation workflow

- Researchers initiate their dataset deposit
- Submissions are reviewed by librarians following a modified CURATE(D) model to ensure that they align with our deposit guidelines
- Librarian will contact the researcher for more information and work together to create/augment files
- On publication, DOIs are minted and flagged for preservation processing
- Uvic Borealis certified by CoreTrustSeal as a Trustworthy Repository

UVic's Data Review Checklist

Federated Research Data Repository (FRDR)

- National generalist repository hosted by the Digital Research Alliance of Canada
- Focuses on housing large datasets
- Works with Globus for seamless file transfer from national clusters, lab machines, and personal computers
- Mints DOIs for each dataset
- Default allocation of 1TB, with possibility to scale up
- Curation services by FRDR team
- Connected to Lunaris data discovery layer

[source](#)



Other Data Repositories

<https://www.re3data.org/>

Creative Commons (CC) Licenses

<https://creativecommons.org/>

- Tells a re-user what they can do with a copyrighted work
- Notes:
 - You must be the copyright holder to assign one of these licenses
 - Licenses can't be revoked once assigned, so pick wisely!
 - (Ask for help if you need!)
 - Very commonly used by data repositories (including Borealis)

Creative Commons Licenses

Abbreviation	Key Feature(s)	What it means
CC BY	By Attribution	Re-users must credit the creator/copyright holder
CC BY-SA	By Attribution, Share-Alike	Re-users must credit the creator/copyright holder; any new material based on this work must be licensed under the same license
CC BY-NC	By Attribution, Non-Commercial	Re-users must credit the creator/copyright holder; the work cannot be used for commercial purposes
CC BY-NC-SA	By Attribution, Non-Commercial, Share-Alike	Re-users must credit the creator/copyright holder; the work cannot be used for commercial purposes; any new material based on this work must be licensed under the same license
CC BY-ND	By Attribution, No Derivatives	Re-users must credit the creator/copyright holder; no derivatives or adaptations of the work are allowed
CC BY-NC-ND	By Attribution, Non-Commercial, No Derivatives	Re-users must credit the creator/copyright holder; the work cannot be used for commercial purposes; no derivatives or adaptations of the work are allowed

Questions?

