

Best practices for managing data in your research.

Isaac Pratt, PhD + Danica Evering, MA

January 18, 2023





McMaster University is located on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the “Dish With One Spoon” wampum agreement.

Laslovarga, “Webster Falls in Winter, Waterdown, Hamilton, Ontario, Canada - Spencer Gorge / Webster's Falls Conservation Area,” 23 January 2011, Wikimedia Commons - https://commons.wikimedia.org/wiki/File:Waterdawn_Webster_Falls_in_Winter8.jpg

Code of Conduct

The Sherman Centre and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

*Please refer to our code of conduct webpage for more information:
scds.ca/events/code-of-conduct/*

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences. In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Certificate Program

The Sherman Centre offers a Certificate of Completion that rewards synchronous participation at 7 workshops. We also offer concentrations in Data Analysis and Visualization, Digital Scholarship, and Research Data Management.

*Learn more about the Certificate Program: <https://scds.ca/certificate-program>
If you would like to be considered for the certificate, verify your participation in this form: <https://u.mcmaster.ca/verification>*

At an unspecified point during the workshop, a code will be read aloud. This is the answer to the third question of the form.

Hello! A bit about us:

We are **Research Data Management Specialists**

Isaac Pratt, PhD

My background is in **Biological Anthropology, Medical Imaging, and Human Anatomy.**

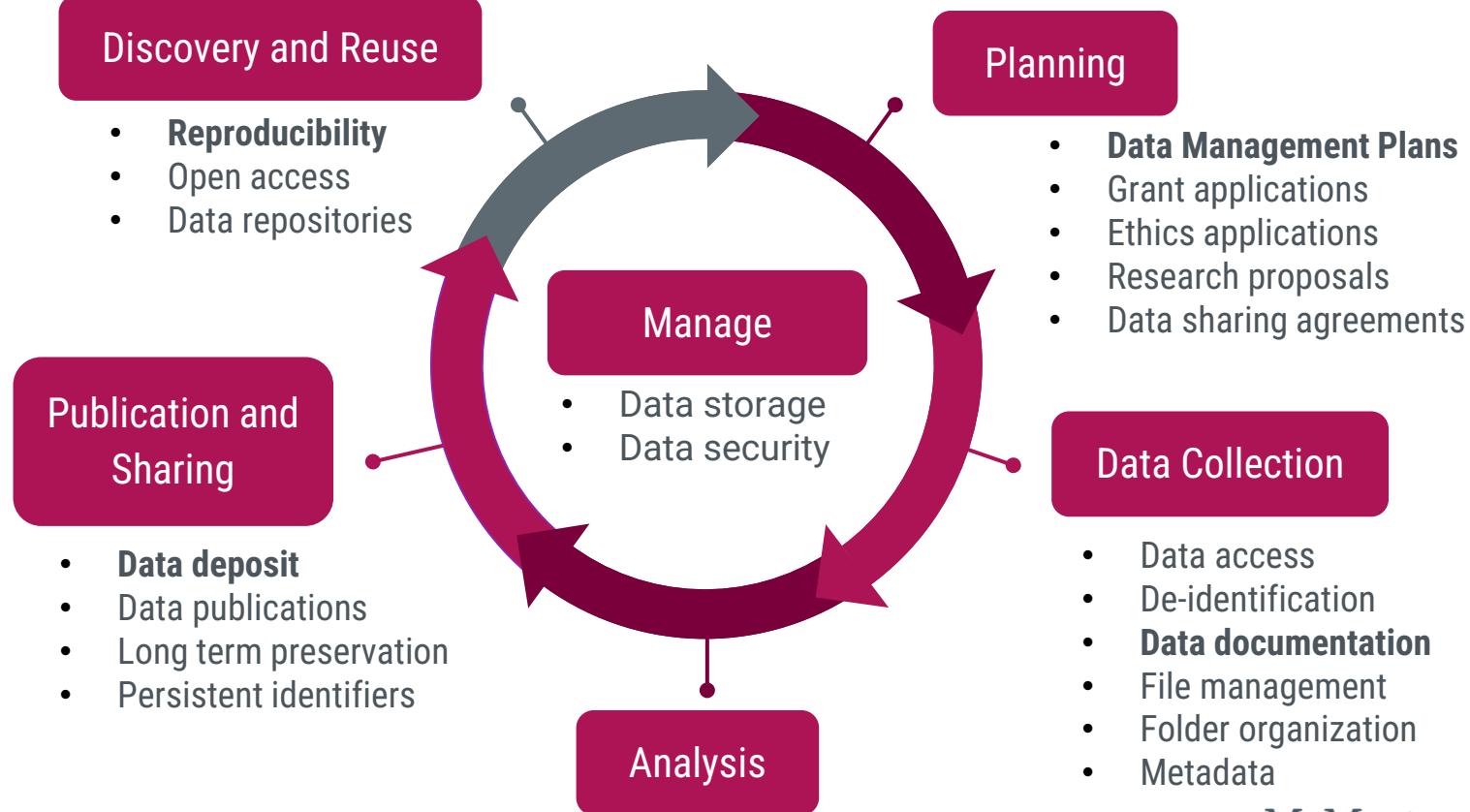
I have a PhD in **Anatomy & Cell Biology** from the University of Saskatchewan.

Danica Evering, MA

My background is in **social practice art, community-based research, communications studies, and medical laboratory healthcare.**

I have an MA in **Media Studies** from Concordia University.

What is Research Data Management anyways?



Is your data vulnerable?

- In a year or two, will you be able to remember all the details of your experiments? Have you recorded all the little details?
- What will happen to your data when you graduate/move/retire?
- How much work would you have to do if you lost your data?
- If you needed to share your data with a collaborator, would they be able to understand it without your help?

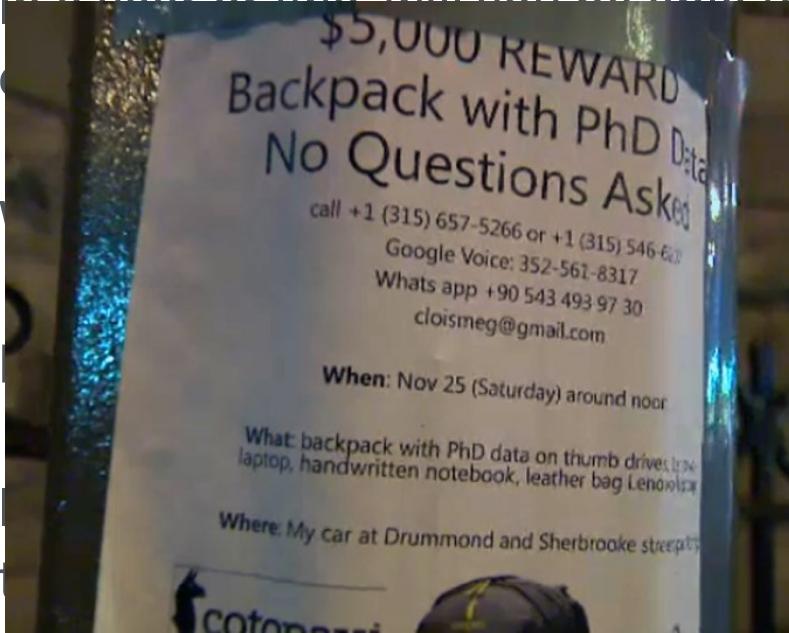
Is your data vulnerable?

- \$5,000 REWARD
Backpack with PhD Data
No Questions Asked

call +1 (315) 657-5266 or +1 (315) 546-6832
Google Voice: 352-561-8317
WhatsApp: +90 543 493 97 30
cloismeg@gmail.com

When: Nov 25 (Saturday) around noon

What: backpack with PhD data on thumb drive, laptop, handwritten notebook, leather bag



FOR MY LOST LAPTOP

I am a Rutgers Chemistry 5th year PhD student. On April 19th afternoon, my LENOVO THINKPAD T420S laptop was stolen from room 203 of Wright-Rieman building. If you stole my laptop and now you are reading this letter, I would like to say that you can keep the computer and I would like to pay you money for my data under D drive. The data is my FIVE-YEAR work. I really need the data under the D drive, there is a folder named RESEARCH, under RESEARCH folder, there is a THESIS folder. I only need that folder for my thesis defense, which is coming very soon. I would like to pay you \$1000 and use whatever way you offer to send you the money. The price is negotiable. My laptop password is 850713zd, my email address is [REDACTED] and phone number is [REDACTED]

[REDACTED] PLEASE contact me and I would appreciate it so so much!!!

Is your data vulnerable?

- In a year, the National Museum of Brazil experienced a massive fire.
 - What were they able to do?
 - How much data was lost?
 - If you run a digital collection, are you able to understand the risks?
- 
- your
?
STEVE MARCUS/REUTERS

Is your data vulnerable?

University of Manitoba Psychology

- In a year or two, will you still be able to understand the details of your experiments? Have you saved the details?
- What will happen to your data if you leave/move/retire?
- How much work would it take to clean up your data?
- If you needed to share your data with another researcher, would they be able to understand it without your help?



Winnipeg Free Press



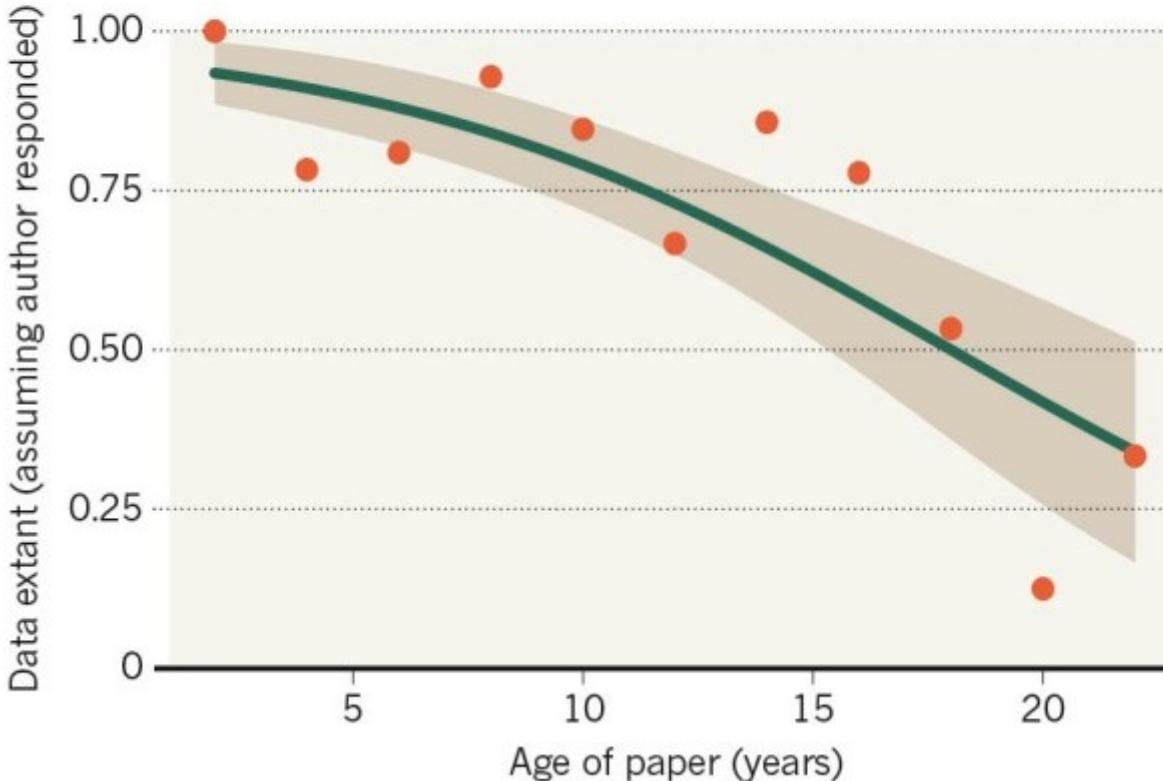
Is your da

MISSING DATA

Vines et al 2014

As research articles age, the odds of their raw data being extant drop dramatically.

- In a year experiment
- What will happen
- How much data
- If you need to understand



f your
e?
ey be able

Let's look at an example:

- “Dave” is a graduate student working in Biomedical Science, focused on x-ray imaging of bone tissue samples. Dave’s data is made up of 3 major components:
 - Image files – x-ray scan images, microscope images
 - Software/hardware configuration files – instrument specific files, scripts, text files
 - Measurement data files – spreadsheet files

Example practices

- Dave's data is stored separately in a few places:
 - Image files are large (2+ TB) and stored on lab computers and a collection of miscellaneous external hard drives accumulated over the years.
 - The other files are smaller (10 GB) and stored on a personal laptop and a cloud storage platform (OwnCloud).
- Data is not consistently documented
- Data is not published or shared outside the research group except by direct request. No time or energy is put into archiving the research data.

What went wrong for Dave

- One of the external drives fails, leading to the loss of some of Dave's data. This data loss is not discovered for several weeks. **There is no back up of this data.**
- This leaves Dave with two choices:
 - Extend his degree while he recollects that data, or
 - Publish what he can, even though the explanatory and statistical power of the study has been reduced



What could Dave do better in the future?

1. Make a plan for your data
2. Organize and document your data consistently
3. Store and back-up your data securely
4. Make sure your data is ready for archival and sharing

- Avoid duplication of research and increases use of existing data
- Save time and resources
- Helps you remember what you did
- Avoid loss of data
- Protect research participants
- Meet funder & journal requirements
- Increase visibility and citation credits

Data Management Planning

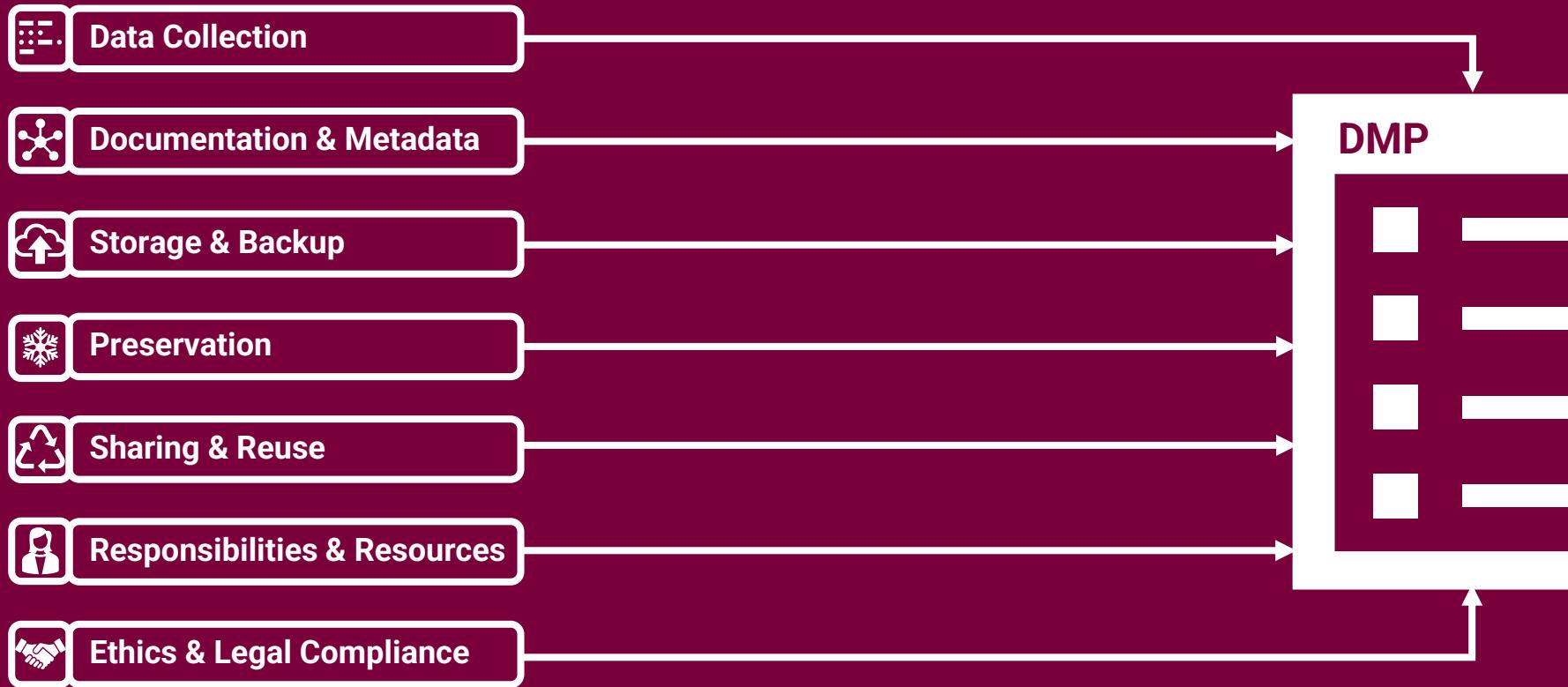
A **Data Management Plan (DMP)** is a living document describing your plan for how you will manage your research data.

Building a DMP is a proactive process that helps you anticipate and identify opportunities and challenges in managing your data before those opportunities and challenges emerge.

A document which speaks to the management of data both **during** the active phases of your research and **after** the completion of the research project.

Some research funders require grant applicants to submit a DMP – NSF, NIH, Wellcome Trust, Tri-Agency (SSHRC, NSERC, CIHR)

What goes in a Data Management Plan?





The screenshot shows the homepage of the DMP Assistant. At the top, there's a navigation bar with links for Home, Public DMPs, DMP Templates, Help, and About. To the right of the navigation is a language selection dropdown. Below the navigation is a large "Sign in" and "Create account" button. To the left of the sign-in form is a "DMP ASSISTANT" logo. The main content area features a "Welcome to DMP Assistant." message. Below this, a paragraph explains what the DMP Assistant is and how it supports researchers. It also mentions that DMPs are one of the foundations of good research data management. A "Getting started:" section lists various resources, including brief guides, primers, tutorial videos, and webinars. At the bottom of the page, there's a footer with logos for Digital Research Alliance of Canada, Alliance de recherche numérique du Canada, DMPonline, CARL ABRC, University of Alberta Library, UC3, and DCC. The footer also includes copyright information and links to Contact us, Terms of use, Privacy statement, and GitHub.

A web-based, bilingual data management planning tool

Available to all researchers in Canada

Walks you through relevant questions for data management

Exportable data management plans

assistant.portagenetwork.ca/



Photo by Crawford Jolly and Clint Adair on Unsplash, edited by Danica Evering.

Organize and document your data

Raw data isn't easy understand and re-use. Data documentation supports **interoperability** and **reusability**.

- If you needed to use data you collected a year ago, how easy would they be to find and use?
- Would you know what every variable is?
- Would you have information about when/where/how the data was collected?
- Have you ever gone to analyze data or publish a paper only to find that some critical piece of information was not recorded, or you don't remember where you wrote it down?

Soil radiometrics: Field and remote data sets for model building and va

242.3KB

Contributors: Cassia Read, David H. Duncan, Chiu Yee Catherine Ho, Matt D. White, Peter A. Ves

Date created: 2017-05-02 09:40 PM | Last Updated: 2018-06-14 11:46 PM

Category: Project

Description: Repository for model training and testing data sets for the article: Read C

White M, Vespa PA. Useful surrogates of soil texture for plant ecologists from airborne &

Ecol Evol. 2017;00:1–10. <https://doi.org/10.1002/ece3.3417>

Wiki

This project is home to the soil data for north-west Victoria, Australia used by Read et al. in their paper "Useful surrogates of soil texture for plant ecologists from airborne gamma-ray detection".

See the respective data set wiki pages for further information on provenance and methods.

Files

Cassia Read, David H Duncan, Chiu Y C Ho, Modified
Matt D White, and Peter A Vesk, "Soil model ...
Radiometrics: Field and Remotely-Sensed
Data Sets for Model Building and Validation,"
OSF, June 15, 2018, osf.io/uac6x.
BRT model building (training) data set

— ↗ OSF Storage (1 United States)

Research Project Management

- **Collaboration:** Google Docs and Microsoft Office let your team work on documents together in real-time, avoiding multiple versions and copies sent by email.
- **Reference Management:** Zotero, Mendeley, or Endnote support collaboration through shared citation libraries.
- **Notetaking software:** Obsidian, Evernote, OneNote, Notion, or an **Electronic Lab Notebook** allow you to create organized, linked notes that you can use to document your research practices
- **Open Science Framework (OSF):** This free open platform for research collaboration lets you manage files, data, code, and protocols in one central location.

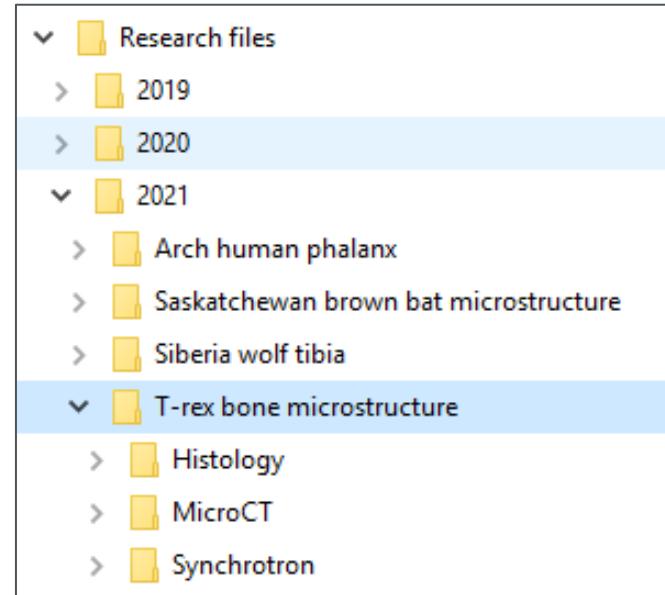
Learn more at rdm.mcmaster.ca/organize

Keeping files organized makes it easier to find things.

The key to organizing files is to make it a habit. Make it easy to know files go.

File organization schemes can include:

- By project
- By researcher
- By experiment type
- By date (often year)
- By some combination of the above
(ie a two level structure of year -> project)



Give your files good names!

Organizing your research files makes your data easier to understand, share, and archive—both now and in the future. A good file organization system should be **descriptive**, **standardized**, and **implemented consistently**.

- **File Naming:** Include date, project name, short description, initials of researcher, version number, and other metadata (like location). Good file names make your files **searchable**.
e.g. 2022_10_01_LakeMercury_TestData1_TM_v3.csv
- **Update your team:** Once you've established a file organization and naming system, describe it in a README file for your research project and make sure everyone on your research team knows the system.

Date: 2022_10_01 (collection date)
Project Name: LakeMercury
Short Description: TestData1
Name: TM (Tracy MacDern)
Version Number: v3

File naming

Do you have files named like this?

Is this a good file name system?

A STORY TOLD IN FILE NAMES:					
Filename	Date Modified	Size	Type		
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file		
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file		
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file		
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file		
data_2010.05.28_huh???.dat	7:20 PM 5/28/2010	30 KB	DAT file		
data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file		
data_2010.05.29_aaarrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file		
data_2010.05.29_#\$@*!&!.dat	2:40 AM 5/29/2010	0 KB	DAT file		
data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file		
data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file		
data_2010.05.29_woohoo!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file		
data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file		
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file		
ThesisOutline.doc	7:26 AM 5/29/2010	38 KB	DOC file		
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file		
JUNK...	2:45 PM 5/29/2010		Folder		
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file		

Project: Kristin's important chemistry project

Date: June 2013-April 2014

Description: Description of my awesome project here

Funder: Department of Energy, grant no: XXXXXX

Contact: Kristin Briney, kristin@myemail.com

ORGANIZATION

All files live in the 'ImportantProject' folder, with content organized into subfolders as follows:

- 'RawData': All raw data goes into this folder, with subfolders organized by date
- 'AnalyzedData': Data analysis files
- 'PaperDrafts': Draft of paper, including text, figures, outlines, reference library, etc.
- 'Documentation': Scanned copies of my written research notes and other research notes
- 'Miscellaneous': Other information that relates to this project

NAMING

Raw data files will be named as follows:

"YYYYMMDD_experiment_sample_ExpNum"
(ex: "20140224_UVVis_KMnO4_2.csv")

STORAGE

Documentation files

- **README:** A simple text document (.txt) that describes project information, folder hierarchy and file organization, description of important file contents.
- **Data Dictionaries:** A document for tabular data that describing names, labels, units, and constraints.
- **Codebooks:** Like data dictionaries but for survey or statistical data—includes the survey layout and structure, and codes for questions and answers.

Build a documentation scheme you will actually use!

The most important aspect of documentation is doing it.

Whatever file naming and organization scheme you choose, make sure it's **descriptive**, use it **consistently** and **document** it (in a `readme.txt` file).

Take advantage of the software that is out there, including note-taking software, reference management software, and collaboration software.

Storing your data safely

1. Avoiding data loss
 - Theft or loss of devices, accidental damage or destruction
2. IT Security
 - Computer viruses, malware, ransomware



Backup Strategies (3-2-1)

A good data storage plan needs to balance accessibility and convenience against security and reliability.

3

Copies of your data (at least!)

Example:

1 copy stored locally on **hard drive** for analysis
1 copy stored on **cloud storage** platform
1 copy stored in a **secure campus drive**

2

Copies are on-hand (easily accessible) on different systems (internal hard drive, cloud storage, etc.)

- a “**production**” (working) copy
- a “**production backup**” copy

1

Copy is in another location (“off-site”) from the others with a **trusted** service provider

Research Data Storage Finder Tool

<http://u.mcmaster.ca/storagefinder>

McMaster RDM Services has a **Data Storage Finder**, an interactive tool to help you find a vetted storage provider depending on risk, volume, and other needs.

This tool also allows you to compare feature sets of selected options.

The screenshot shows the 'Research Data Storage Finder Tool' interface. It consists of two main sections: 'Step 1' on the left and 'Step 2' on the right.

Step 1: Answer these questions to narrow down storage provider options.

CLEAR ANSWERS

1. What risk level is your data?

- Low
- Medium
- High

2. What type of data storage are you looking for?

Step 2: Select data storage providers you would like to compare

SELECT ALL **CLEAR SELECTIONS**

Compute Canada	Compute Canada NextCloud	Dataverse
Advanced research computing systems, storage and software	Advanced research computing File hosting services	Store, share, publish and discover research data
FRDR	Github	MacDrive
Find and Share Canadian Research Data	Distributed version control system for software code	File Synchronization and Sharing solution
MacDrop	McMaster based custom solution	OSF
Web service to store and		Open platform for



Encrypt sensitive data:

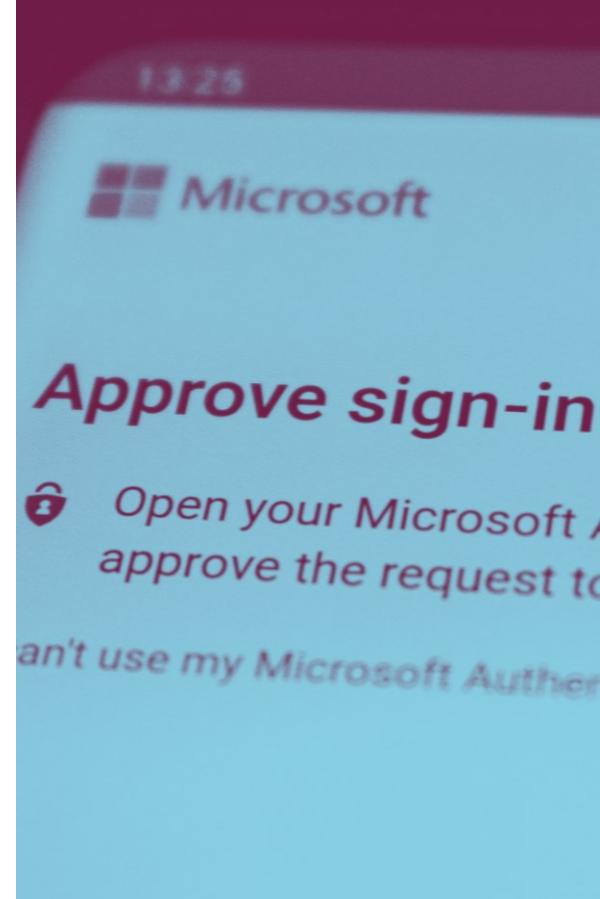
Sensitive data is any data that would cause harm if released openly. This includes personally identifiable information and personal health information.

- Use “**Full disk encryption**” if you are using a personal computer or laptop. This is called FileVault on Mac OS and “**device encryption**” or Bitlocker on Windows.
- Encrypt individual files in Microsoft Office using the “Protect Document” function.
- Other files can be encrypted using Disk Utility on Mac OS or with a third-party tool like VeraCrypt.

How should I protect my data?

Enable Multi-Factor Authentication (MFA)

- Also known as 2 Factor Authentication (2FA)
- Requires more than one code or 'Factor' to login – typically 2 factors: password and a security code sent to your phone number or generated by a linked authenticator app
- Many other web services (Gmail, Dropbox, etc) provide MFA



Password Best Practices

Make sure your online information is secure by ensuring your password is:

- ✓ **Strong:** Make a strong password by combining a series of numbers, letters, and symbols into a long series of words. Try to combine them into something memorable – like L1br@ryt1pS.
- ✓ **Unique:** Use a different password for every website/service
- ✓ **Secret:** Never share your passwords with anybody, even if you trust them. Keep your passwords secret by storing them only in your head. Never send them in an email!
- ✓ **Fresh:** Change your passwords at least once every three months
- ✓ **Devices:** Use a strong password on your computer and phone, too



Tip: Remembering multiple passwords can be difficult. Use a trusted password manager to keep track of your passwords for you. Some examples are BitWarden and 1Password.



Common Password Mistakes

Here are some common mistakes people make when creating passwords:

Service	Email	Password
MacID	example@mcmaster.ca	football
DropBox	example@mcmaster.ca	football
Google	test.example@gmail.com	football
Instagram	test.example@gmail.com	football1
ORCID	example@mcmaster.ca	football
Zotero	example@mcmaster.ca	football

Written Down: Passwords written down on a piece of paper or stored in plain text on a computer may be stolen by somebody with malicious intent or easily lost.

Too Simple: Simple or common passwords are easy to guess or brute force.
Examples: *apple, rowboat, bumblebee, blizzard, password*

Same Password: Using the same password for multiple websites is like having one key for multiple locks; if it's stolen, the thief can open them all.

Have you made any of these mistakes before?

Publishing Data

What do you plan to do with your data once it's been published? How will you ensure that your data remains accessible (to you and others) long-term?

Consider the advantages of publishing your datasets in an online repository for preservation and sharing.

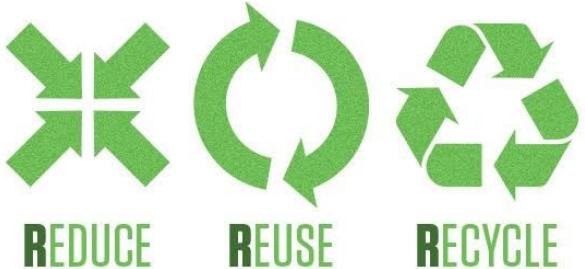
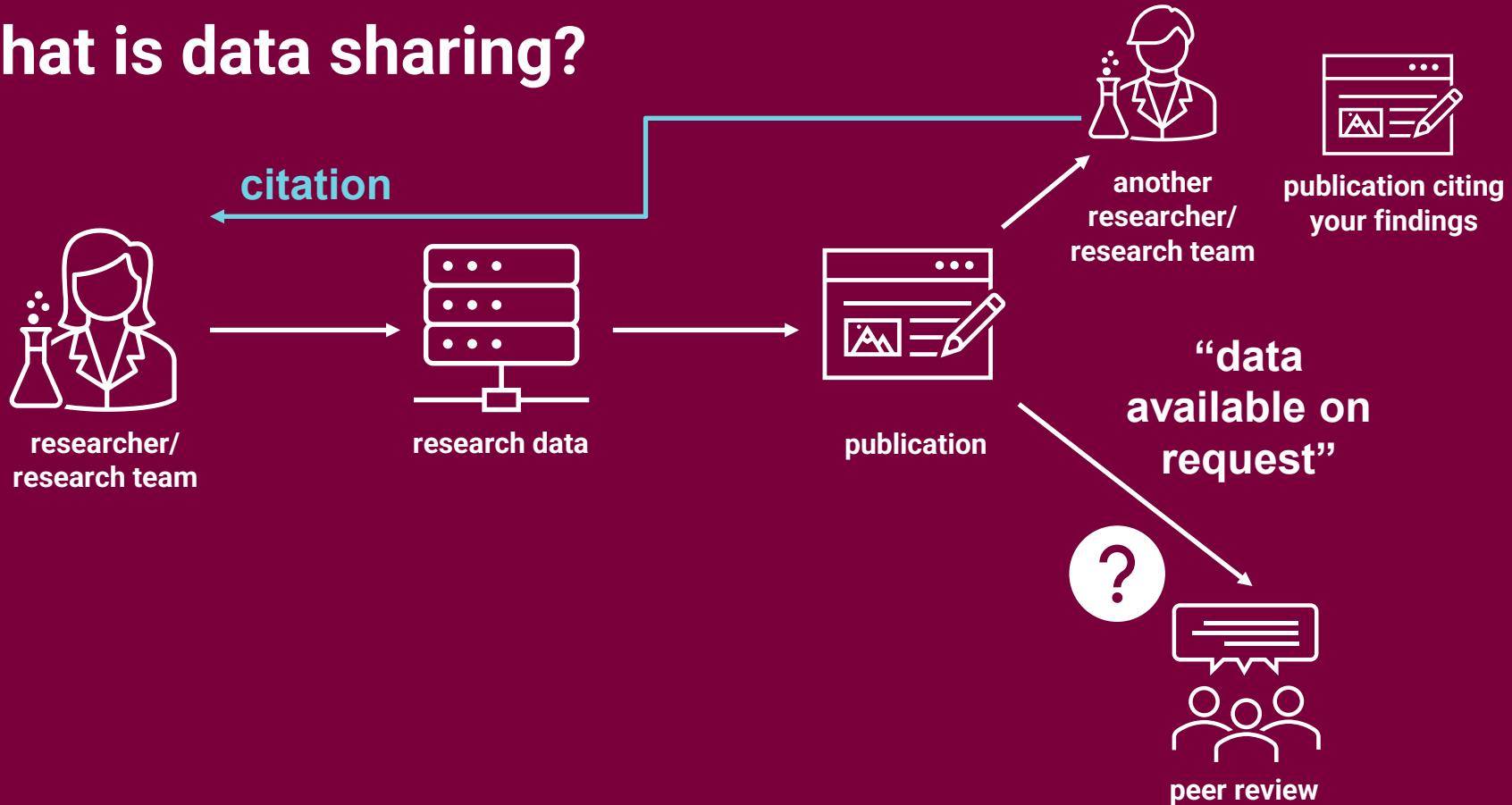


Photo by Lars Kienle on Unsplash.



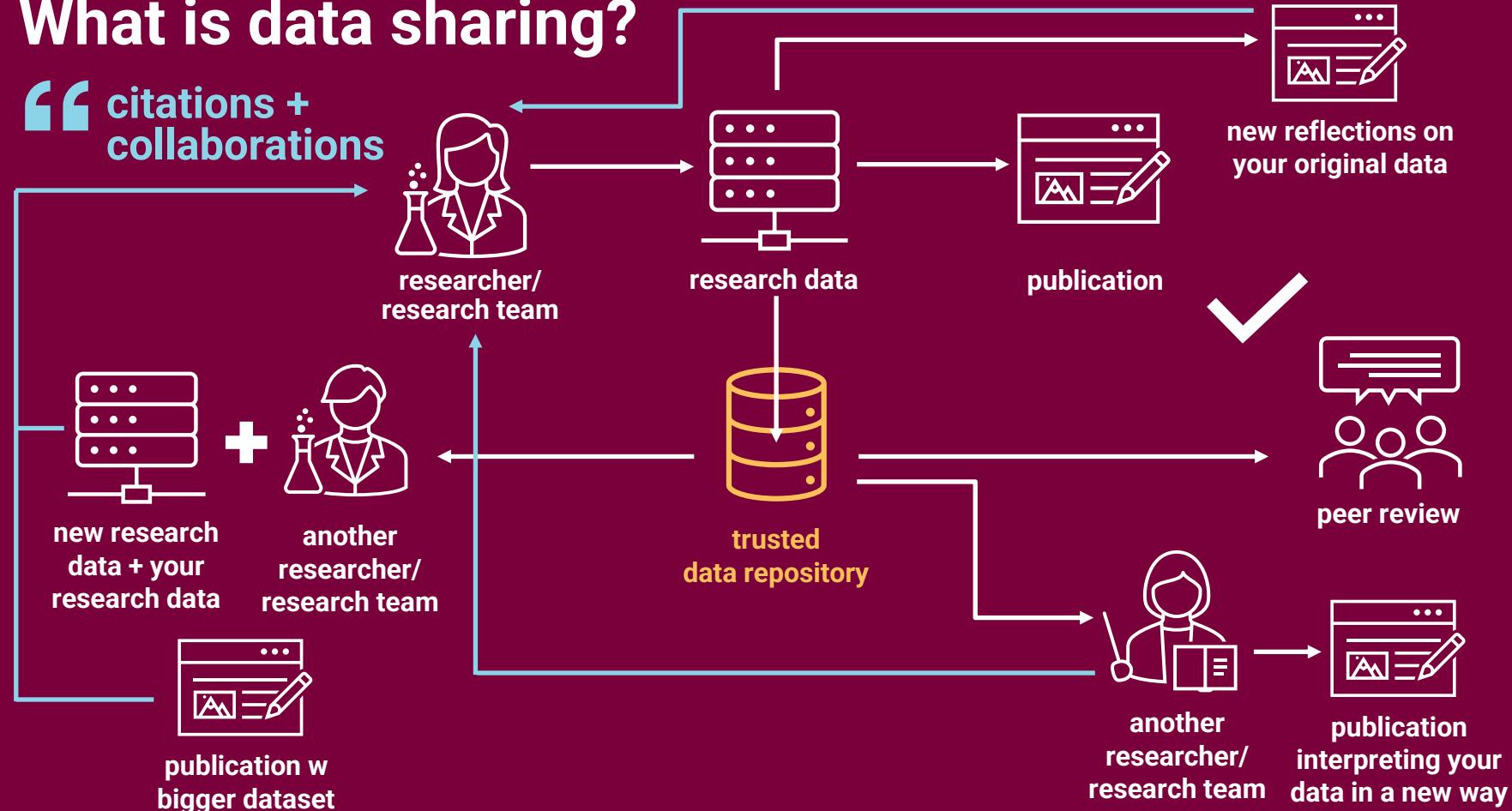
Open Research
open data
open-source software
open access
open government

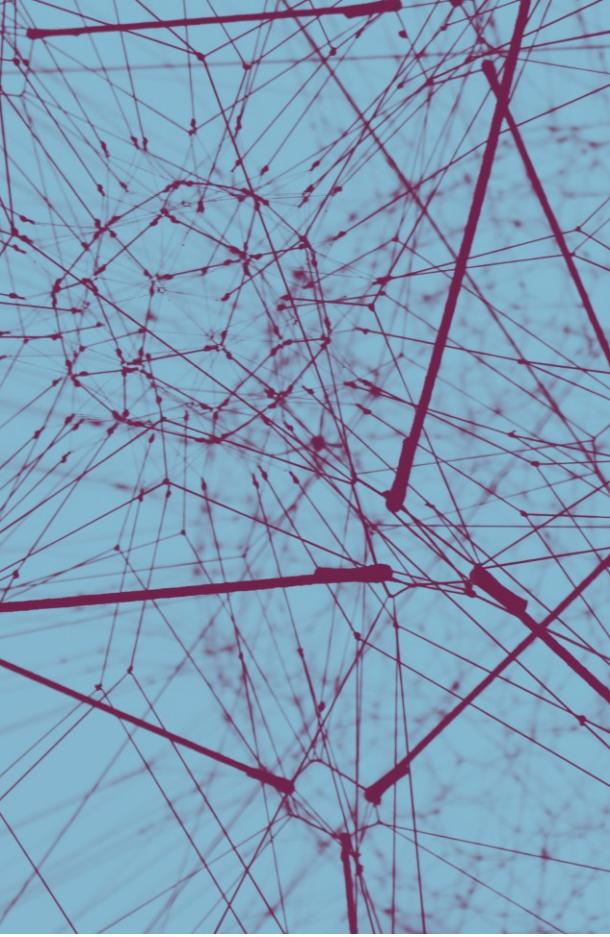
What is data sharing?



What is data sharing?

citations +
collaborations





Why share data? Citation Impact

Studies show that **publications with open data are cited more.**

- Publications in PLOS and BMC journals with open data have up to 25% higher citation impact compared to those that don't share data.
 - Collavazi et al, 2020 PLOSOne The citation advantage of linking publications to research data <https://doi.org/10.1371/journal.pone.0230416>
- Publications of gene expression microarray data have higher citation impact when the data is shared.
 - Piwowar & Vision, 2013 PeerJ Data reuse and the open data citation advantage <https://doi.org/10.7717/peerj.175>



Why share data? Journal and Publisher Requirements

Many journals are starting to require data sharing or at least **data availability statements**, including:

- PLOS <https://journals.plos.org/plosone/s/data-availability>
- Nature <https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards>
- NEJM <https://www.nejm.org/about-nejm/editorial-policies>

Journals with higher Impact Factors are more likely to have data sharing policies.



Why share data? Tri-Agency Requirements



Government
of Canada

Gouvernement
du Canada

Search Canada.ca

MENU ▾

Home > [Interagency research funding](#) > [Policies and Guidelines](#) > [Research Data Management](#)

Research Data Management

Tri-Agency Statement of
Principles on Digital Data
Management

Open Letter

Tri-Agency Research Data Management Policy

Public Consultation
Summary

Frequently Asked
Questions

Tri-Agency Research Data Management Policy

1. Preamble

The [Canadian Institutes of Health Research \(CIHR\)](#), the [Natural Sciences and Engineering Research Council of Canada \(NSERC\)](#), and the [Social Sciences and Humanities Research Council of Canada \(SSHRC\)](#) (the agencies) are federal granting agencies that promote support research, research training, knowledge transfer and innovation within Canada.

The agencies expect the research they fund to be conducted to the highest professorial and disciplinary standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good use of public funds, experiments and studies are replicable, and research results are as accessible as possible. Research data management (RDM) is a necessary part of research excellence.

The agencies believe that research data collected through the use of public funds should be responsibly and securely managed and be, where ethical, legal and commercial

“Grant recipients are required to deposit into a digital repository all digital research data, metadata and code... in journal publications and pre-prints.”

CIHR currently requires researchers to “deposit bioinformatics, atomic, and molecular coordinate data into the appropriate public database.”

SSHRC requires researchers to “make available for use by others all research data collected with the use of SSHRC funds”

See the [Tri-Agency Data Management Policy](#) for details.

Open Access (OA) publishing

Tri-Agency funded research *must* be published open access. We encourage all research to be published open access when possible!

Online Repositories

- Final manuscripts can be deposited in an institutional or disciplinary repository (such as [arXiv.org](https://arxiv.org))
- Researcher is responsible to navigate copyright requirements of the journal

Journals

- Journal provides open access to the article (within 12 months)
- Most journals will charge open access fees

Persistent Identifiers help keep track of everything and make research findable.

Persistent Identifiers (PIPs) are unique links that will never expire.

- Digital Object Identifiers (DOIs): Can be minted for publications, datasets, lab equipment, research software and more.
- Open Researcher & Contributor ID (ORCID): Unique identifier for researchers. Distinguish yourself from scholars with the same name; connect your datasets, code, and publications.

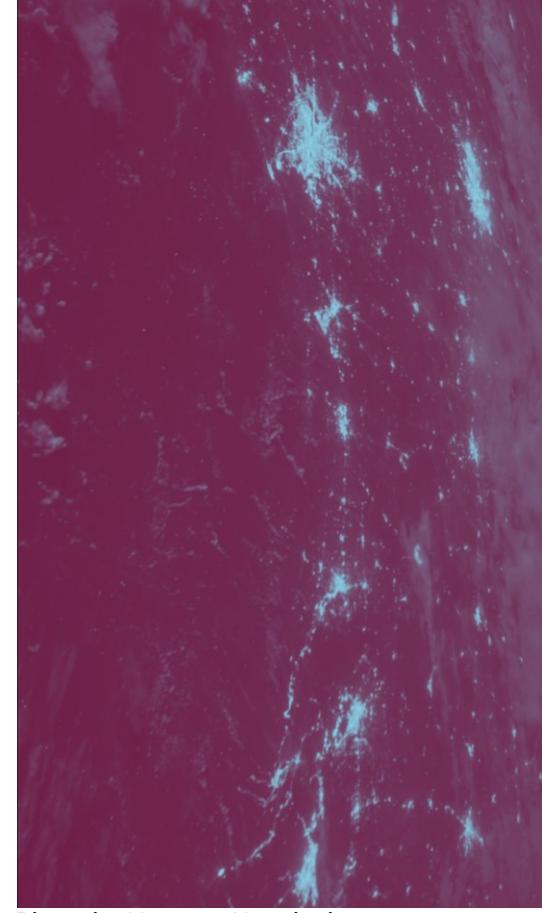


Photo by Nasa on Unsplash.

Sustainable File Formats

Have you ever saved data on a CD, DVD, or BluRay? How about a zip disk or HD DVD? Do you use Google Docs or Prezi where all your documents are stored online on a proprietary platform in a proprietary format? What would you do if that platform closed down?

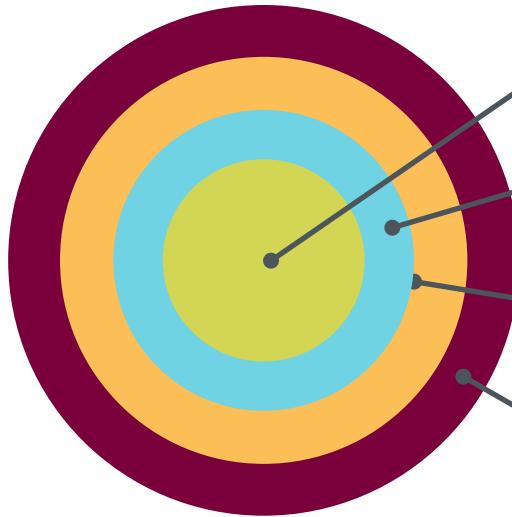
Other researchers may not have access to any proprietary software you use, so data and metadata should ideally be stored in **sustainable formats**. Look for formats that are:

- Standardized
- Well documented
- In common usage
- Uncompressed

Research instrument files may be manufacturer specific and should be converted to a sustainable format when possible. See
[https://site.uit.no/dataverseno/deposit/prepare/
#what-are-preferred-file-formats](https://site.uit.no/dataverseno/deposit/prepare/#what-are-preferred-file-formats)



Datasets as a digital objects.



Research output (data/code): The data is surrounded by layers of information to make it FAIR



Identifiers: Persistent Unique Identifiers such as DOIs and ORCiDs help find, track, and cite data



Standards: Open standard file formats help others access and reuse data



Metadata: Rich metadata and data documentation helps others find and understand datasets



Do I need a license for my data?

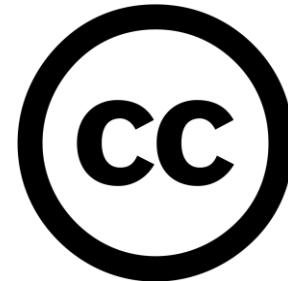
If you don't have a license for your data or code, it falls under the default copyright laws. This means nobody else can copy, distribute, or modify your work without being at risk.

Not having an explicit license restricts others from using your code or data and causes confusion.

Data Licensing

- **Creative Commons (CC)** - (creativecommons.org)
 - CC0 – public domain dedication
 - CC-BY – require attribution
 - There are further restrictions that can be added such as NC
- **Open Data Commons** (opendatacommons.org)
 - Similar licenses to CC but built for data
 - PDDL - Public Domain Dedication and License
 - ODC-By – require attribution
 - ODbL – attribution and share alike
- **Traditional Knowledge (TK)**: In addition to the First Nations Information Governance Centre's OCAP® training, communities may also use TK licenses - localcontexts.org/licenses

Images from Sq'ewlets, "Traditional Knowledge Labels," http://digitalsqewlets.ca/traditional-knowledge_connaisances_traditionnelles-eng.php and **Creative Commons**, fixed by [Quibik](#).



In our Stó:lō culture, certain types of knowledge are restricted in some way. This knowledge is considered sacred, secret, potent and/or private, and only certain people or families can and should have access to them. We call this xa:xá in our language. This label indicates that there is additional knowledge about a certain subject that cannot be shared on the website.

Community Norms

For data there are also **community norms**.
Dataverse and Open Data Commons
community norms include:

- Share your work too
- Credit and Cite datasets you use
- Maintain anonymity of human research participants
- Encourage others to reuse data
- Use open formats
- Don't use Digital Rights Management (DRM)

<https://dataverse.org/best-practices/dataverse-community-norms>
<https://opendatacommons.org/norms/>

Community Meeting 2022
June 14, 15, and 16



[Speakers and Chairs](#) [Registration](#)



#Dataverse2022

[Bonnie Healy's](#) professional background is multi-faceted, having worked in numerous health capacities at local, national, and international levels. Actively involved in various fields of knowledge, Bonnie's passion is to support communities in information data research methodologies. Bonnie's background in First Nations information systems gives her a strong passion for using data as a tool for

The annual Dataverse Community Meeting is an opportunity to build, grow, and enrich the global community. Like the open-source Dataverse product itself, the activities of the Dataverse Community Meetings are community-driven. Over three days of presentations, workshops, and working group meetings we aim to promote and learn about behavioral and technical solutions and standards for curating, sharing, and preserving data that can be discovered and reused across disciplines to reproduce and advance research.

The Dataverse Community Meeting is hosted by Harvard's [Institute for Quantitative Social Science](#). Learn more about The Dataverse Project at our [dataverse.org](#) site.



Ok, so where do I put everything?

A **data repository** is a web platform and storage space for researchers to deposit data sets associated with their research.

Repositories provide:

- long-term storage and access to research data beyond the life of a grant, research project, or individual careers.
- Discoverability and findability for datasets through features like indexing and DOIs.
- Easy-to-use shared platforms made for research.

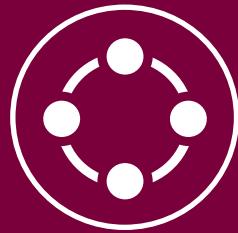
Data Repository Decision-Making

Publishing data in a recognized data repository is the best way to share data. There are thousands of data repositories. [MacSphere is our institutional repository for publications.](#)



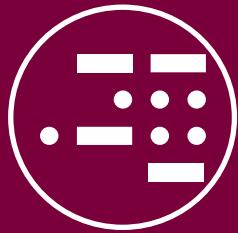
Domain Specific Repositories

Focus on certain types of data such as genomic information or astronomical information.



General Repositories

Accept broader types of research data. ex. *McMaster Dataverse* (part of *Borealis*) and *Canada's Federated Research Data Repository (FRDR)*.



Code Repositories

There are also code-specific repositories like Github, Gitlab, BitBucket, SourceForge



Repository Finder

This tool by DataCite can help you find an appropriate repository to deposit your research data: <https://repositoryfinder.datacite.org>



Ok, so where do I put everything?

MacSphere

<https://macsphere.mcmaster.ca>

- Institutional repository for **scholarly documents (not data)**.
- A home for all research documents, including publications, presentations, conference proceedings, theses, reports, etc.
- When you graduate you will upload your thesis here.
- Link your dataset DOI!



McMaster Dataverse

<https://borealisdata.ca/dataverse/mcmaster>

- McMaster's Institutional Data Repository is a home for all research data originating from McMaster researchers.
- Provides basic data curation services
- Data is stewarded by professionals at McMaster
- Choose whether to share data openly or through an application process

The screenshot shows the McMaster University Dataverse homepage. At the top, there is a navigation bar with links for Search, User Guide, Support, English, and Log In. The main header features the borealis logo, the McMaster University crest, and the text "McMaster University Dataverse (McMaster University)" and "McMaster RDM Services". Below the header, there is a breadcrumb trail showing "Borealis >". On the right side, there are "Contact" and "Share" buttons. The central content area is titled "About Dataverse" and provides information about the repository's purpose and services. It includes a search bar, an advanced search link, and a "Add Data" button. A sidebar on the left lists categories: "Dataverses (19)", "Datasets (64)", and "Files (783)". The main content area displays a list of 83 results, with the first item being "Long-term Rocky Tidal Community Data from Discovery Bay, Jamaica" by Jurek, Kolasa, dated Jun 30, 2022. The list also includes entries for "Nurr1 is not an essential regulator of BDNF in mouse cortical neurons" and "The rocky supraltidal zone is home to a wide diversity of life. Understanding how individual species persist, the interactions between species in a community, and how the abiotic environment can shape these interactions requires detailed study. Over nearly three decades, Professor...".

Federated Research Data Repository (FRDR)

<https://www.frdr-dfdr.ca/repo/>

- Available to any researcher affiliated with a Canadian institution
- Built for large (1 TB+) datasets
- Datasets are actively curated by professional staff at FRDR
- Datasets must be open access but can be embargoed for a one-year period



FRDR Zero knowledge encryption for sensitive data

- FRDR is working on a pilot project to add optional zero-knowledge encryption to the repository.
- “**Zero-knowledge encryption**” means that FRDR will never be able to access your data. All datasets are encrypted, and their keys stored in a separate researcher-managed platform. This allows you to deposit your data in a trusted repository for archival but maintain complete control over access to the data.
- Interested in piloting this service? Email us: rdm@mcmaster.ca



Top 4 ideas for improving your research data management:



Make a **plan** for data management.



Create a **file organization scheme** (and use it).



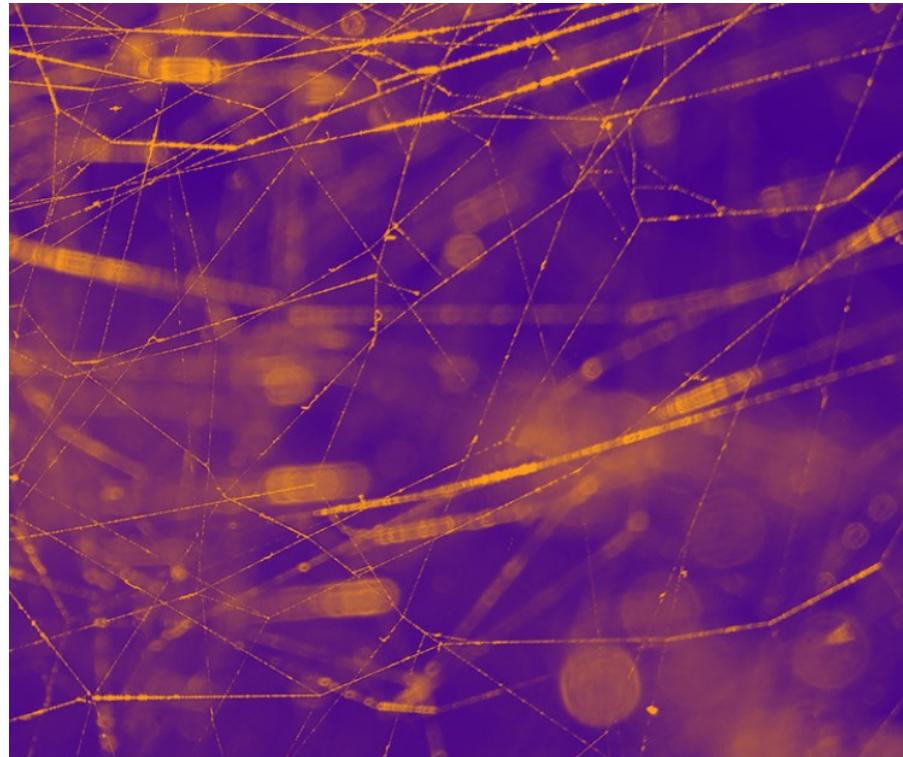
Ensure your data is safely **stored** and backed up.



Share your data openly.

RDM Community of Practice

- Monthly meetings of people interested in RDM at McMaster –
Thurs. Jan. 26 – 11 AM!
- Guest presentation by Dr. Antonio Paez, Earth Environment and Society on Reproducible Research in Graduate Studies.
- Connect with other researchers practicing RDM across the university!
- <https://u.mcmaster.ca/rdm-community>





January 25 | 10:30-11:30am
Virtual Workshop

Depositing & Sharing Data Online with McMaster Dataverse

u.mcmaster.ca/scds-events

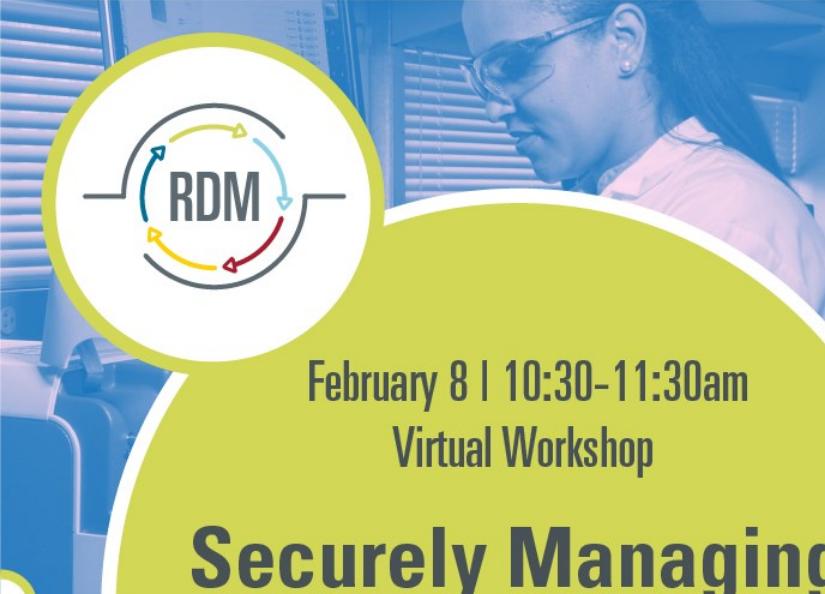


SCDS

Library

McMaster
University





RDM

February 8 | 10:30-11:30am
Virtual Workshop

Securely Managing and Publishing Sensitive Data

u.mcmaster.ca/scds-events



SCDS

Library

McMaster
University





February 15 | 10:30-11:30am
Hybrid Workshop

Persistent Identifiers For All

u.mcmaster.ca/scds-events



SCDS

Library

McMaster
University





RDM

March 15 | 10:30-11:30am
Virtual Workshop

Before You Dig: Finding and Reusing Datasets

u.mcmaster.ca/scds-events



SCDS

Library

McMaster
University





RDM

April 5 | 10:30-11:30am
Hybrid Workshop

Qualitative Data: Practices for RDM Planning & Sharing

u.mcmaster.ca/scds-events

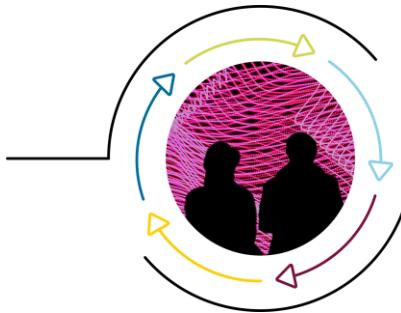


SCDS

Library

McMaster
University





Research Data Management Services

McMaster RDM webpage:

rdm.mcmaster.ca

Contact RDM services at:

rdm@mcmaster.ca

Upcoming RDM webinars:

rdm.mcmaster.ca/events

Recorded RDM webinars:

u.mcmaster.ca/learn-rdm

Make an appointment with a Research Data Management Specialist:
u.mcmaster.ca/rdm-appointments