

September 27, 2023 | 10:30-11:30am  
Virtual Workshop

# Best Practices for Managing Data in Your Research

[u.mcmaster.ca/scds-events](https://u.mcmaster.ca/scds-events)



**SCDS**

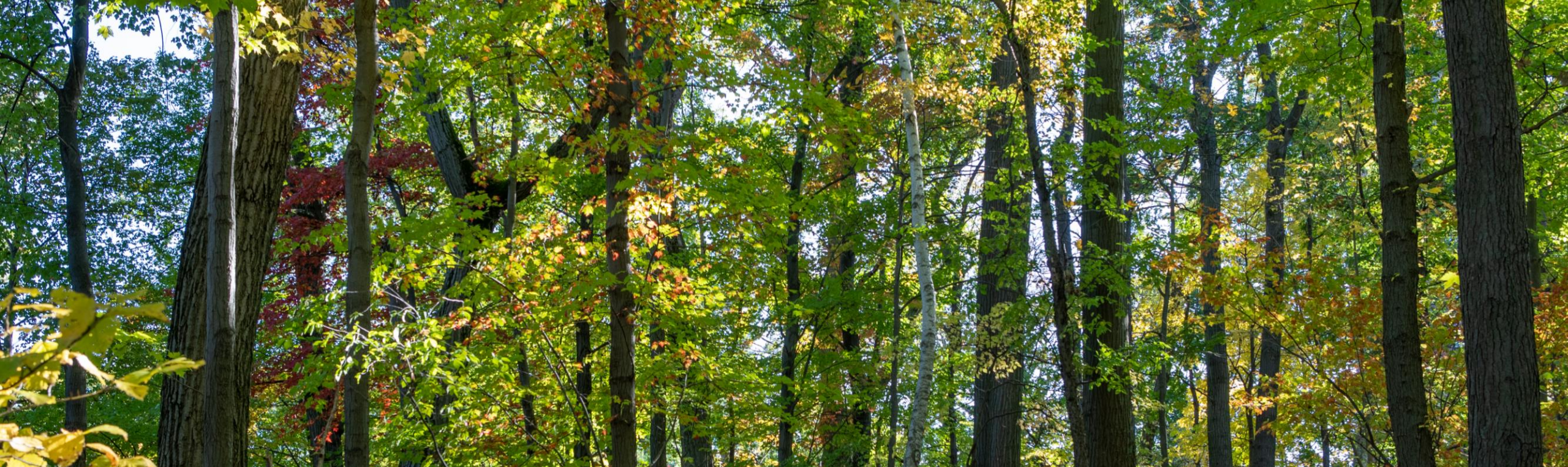
Library



# Best Practices for Managing Data in Your Research

Isaac Pratt, PhD + Danica Evering, MA  
Research Data Management Specialists  
rdm@mcmaster.ca

Research Data Management Workshop Series  
September 27, 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.

Georgia Kirkos, “Cootes Trail,” October 29, 2021, McMaster University, Hamilton, Ontario, Canada

<https://brand-resources.mcmaster.ca/asset-bank/action/viewAsset?id=40841&index=14&total=34&view=viewSearchItem>

# 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/](https://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 Attendance 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>

Verify your participation at a session: <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.

# Research Data Management Workshops

Register for upcoming RDM events: <https://rdm.mcmaster.ca/events>

**Sept. 28:** “RDM Community of Practice - Data Management Plan Roundtable”

**Oct. 25:** “Data Management Plans and Intro to DMP Assistant”

**Nov. 29:** “Depositing & Sharing Data Online with McMaster Dataverse”

**Feb. 14:** “Storage Scores: Store & Back Up Data at McMaster”

**Mar. 20:** “How to Implement Encryption to Protect Your Research Data”

**Apr. 17:** “Sensitive Data Management”

**May 14:** “Data Management Plan (DMP) Bootcamp”

**Jun. 18:** “Data Deposit Bootcamp”

**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.

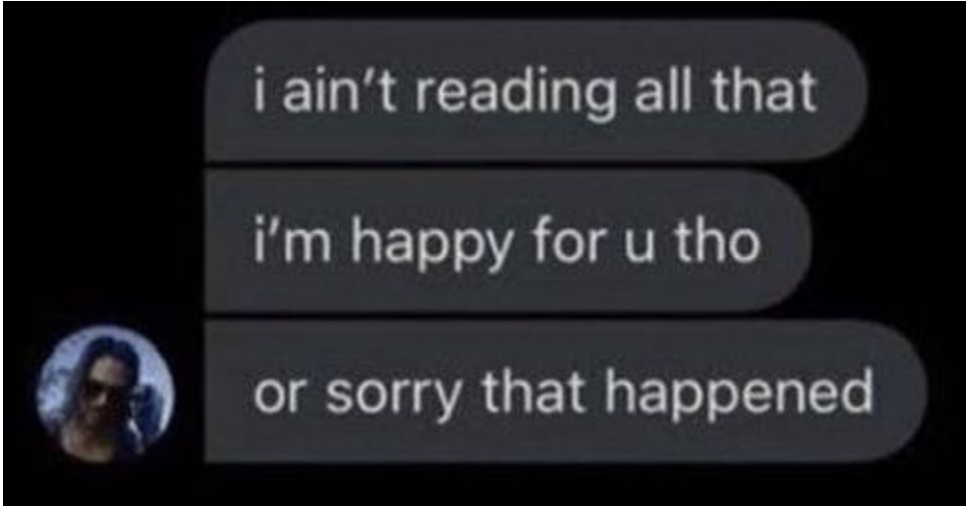
**Danica Evering, MA**

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.



# Outline

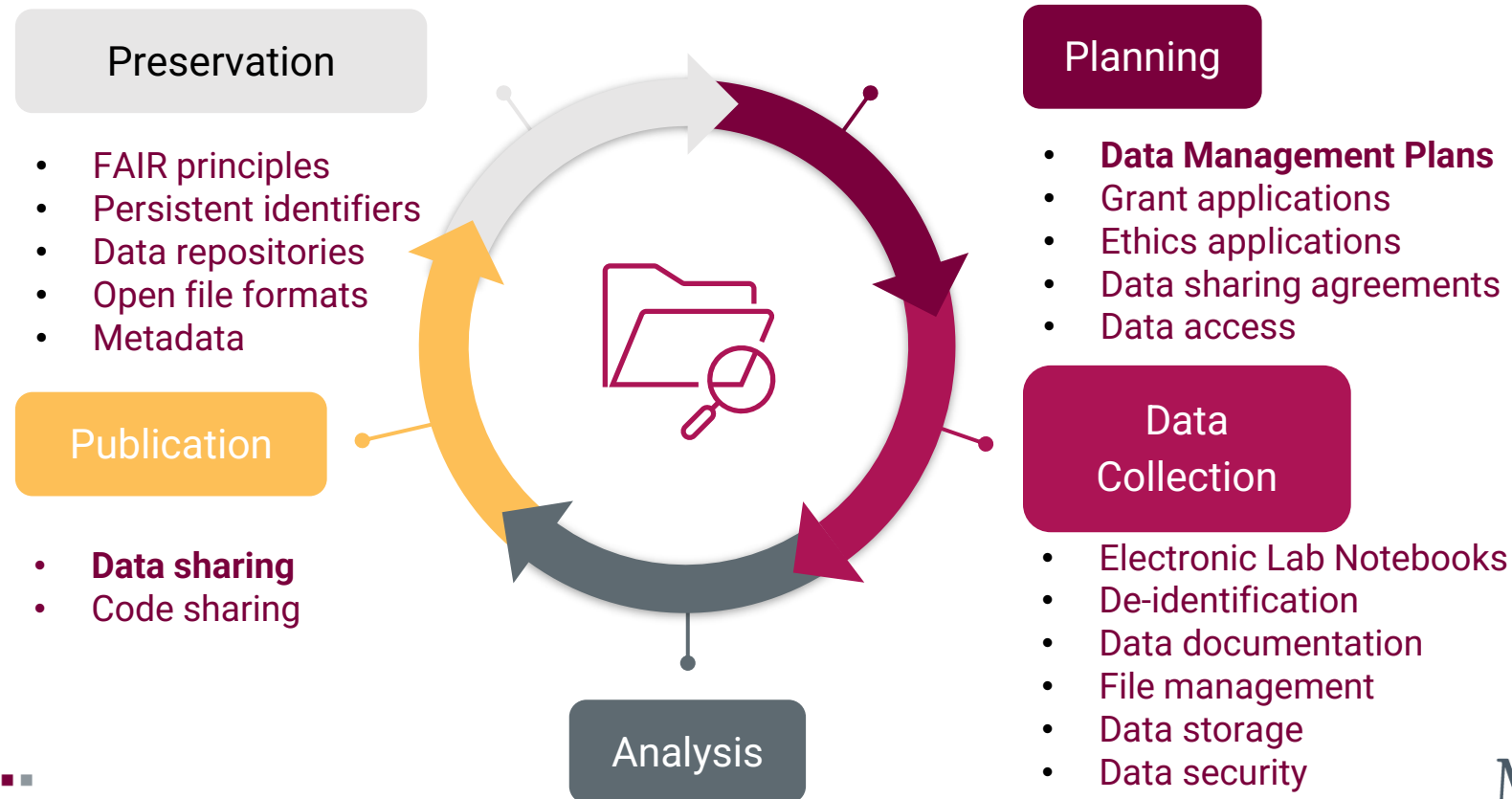


## TL;DR – ~~Best~~ Good Enough Practices for Managing Data in Your Research

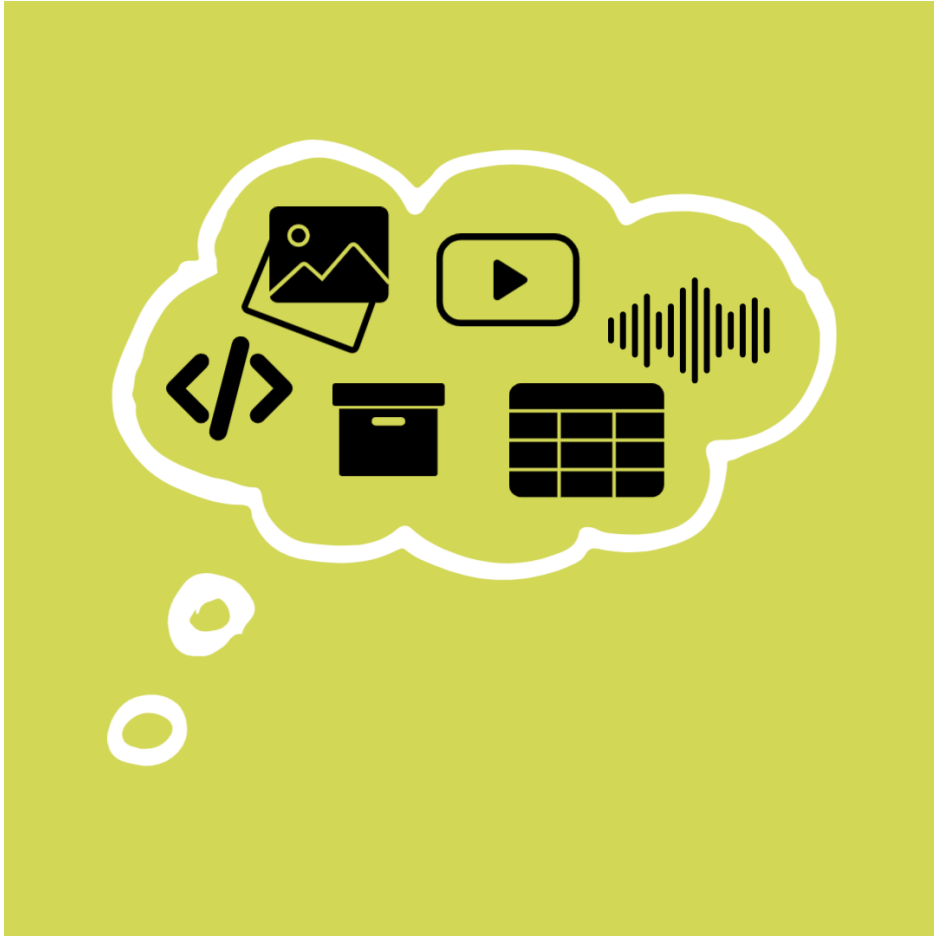
- Everyone has data even (if you don't think you do!)
- Don't be like Dave
  1. Make a plan for your data
  2. Organize and document your data consistently
  3. Store and back up your data securely
  4. **Make sure your data is ready for archival and sharing**

# What is **Research Data Management**?

Research Data Management is a suite of connected processes and practices applied throughout the research lifecycle, as data are **planned** for, **collected**, **organized**, **documented**, **stored**, **preserved**, **shared**, and **reused**, in support of analysis, research, creative works, and dissemination that benefit society.



# *Do I even have research data to manage? (You do!)*



“Data may be in any format or medium taking the form of text, numbers, symbols, images, films, video, sound recordings, pictorial reproductions, drawings, designs or other graphical representations, procedural manuals, forms, diagrams, workflows, equipment descriptions, data files, data processing algorithms, software, programming languages, code, or statistical records.”

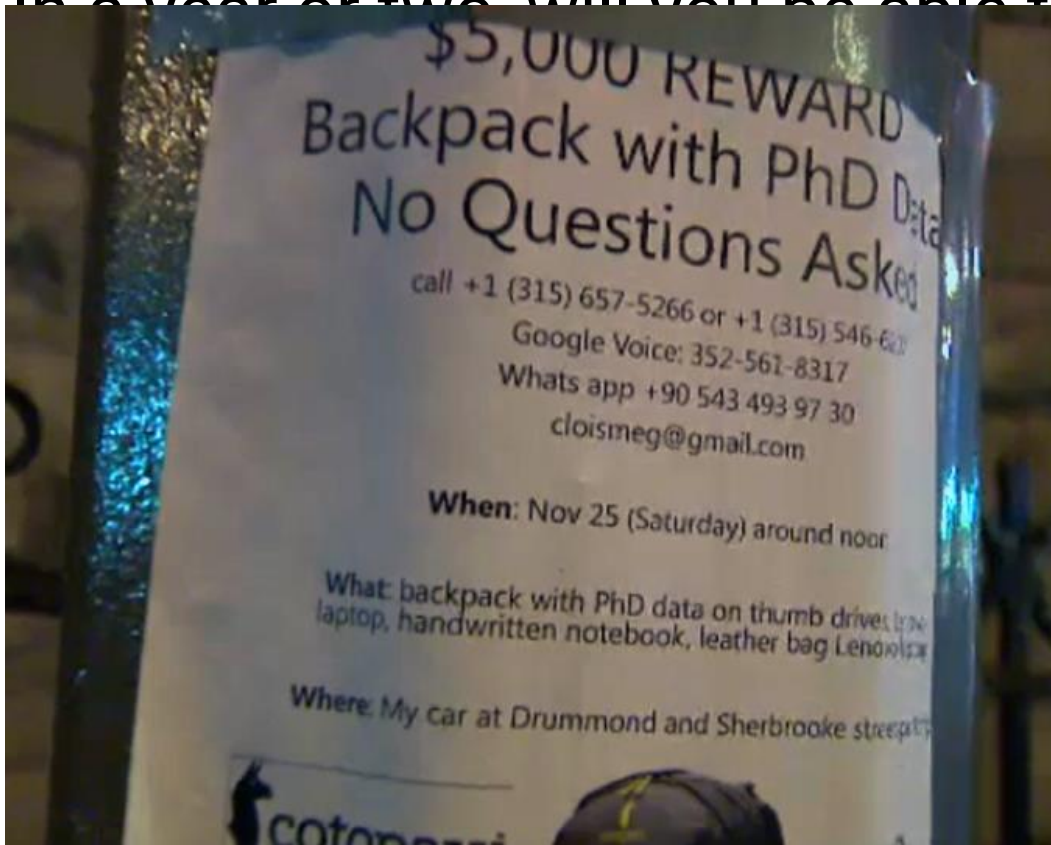
Innovation, Science and Economic Development Canada. “Frequently Asked Questions -Tri-Agency Research Data Management Policy.” Government of Canada. Innovation, Science and Economic Development Canada, October 19, 2021.  
[https://science.gc.ca/eic/site/063.nsf/eng/h\\_97609.html#1a](https://science.gc.ca/eic/site/063.nsf/eng/h_97609.html#1a)

## *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?

In a year or two, will you be able to remember all the details of your



## FOR MY LOST LAPTOP

I am a Rutgers Chemistry 5<sup>th</sup> year PhD student. On April 19<sup>th</sup> 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]. PLEASE contact me and I would appreciate it so so much!!!

# Is your data vulnerable?

- In a year of experience
- What would you do?
- How many people would you be able to

National Museum of Brazil



your  
e?  
they be

# Is your data vulnerable?

University of Manitoba Psychology

- In a year or two, will you still be able to access your experiments? Have you backed up your data?
- What will happen if you leave the university or move/retire?
- How much work will you have to do to migrate your data?
- If you needed to sell your data, would they be able to understand it?



the details of your experiments?

move/retire?

your data?

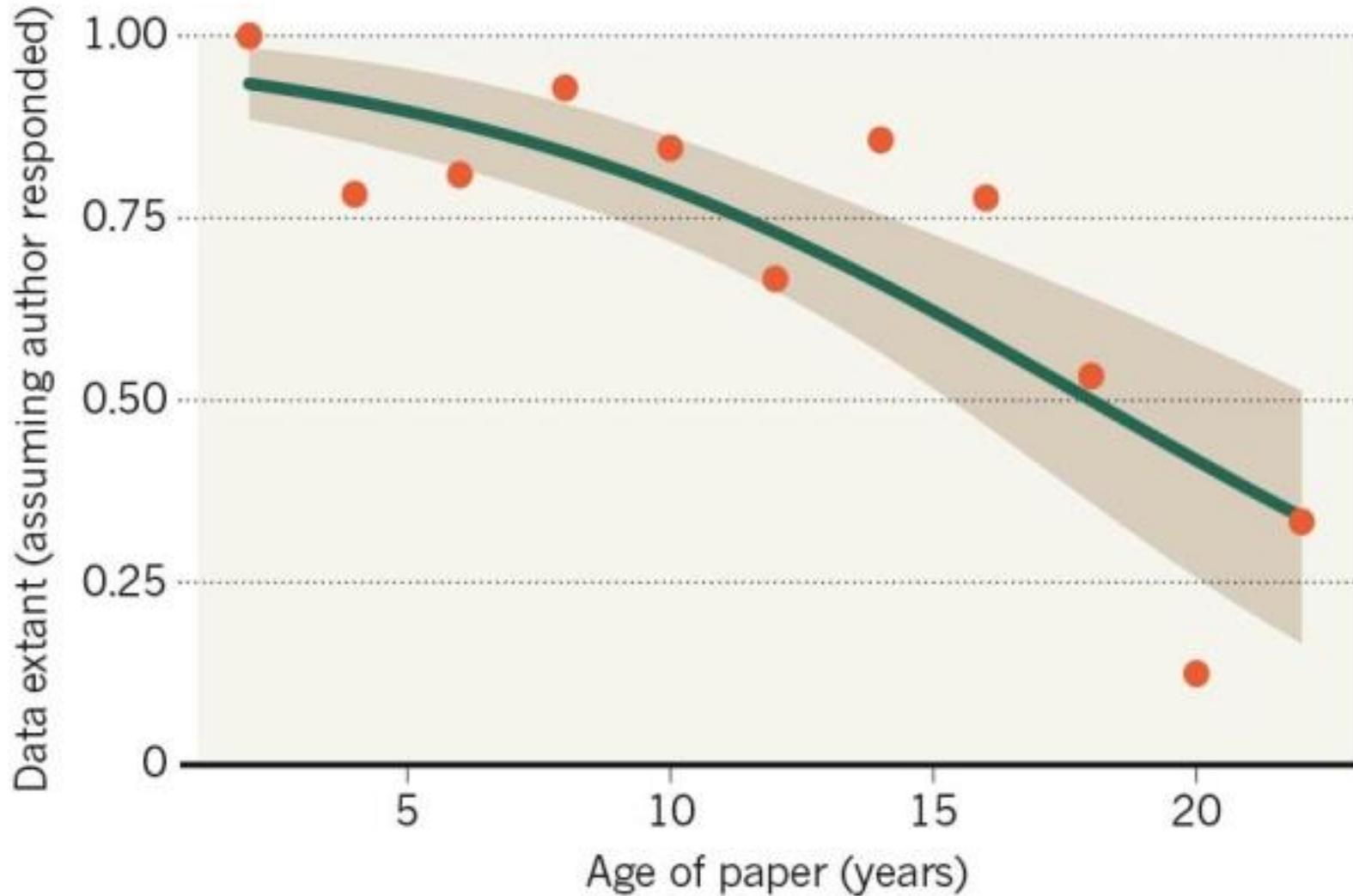
if you needed to sell your data, would they be able to understand it?

# MISSING DATA

Vines et al 2014

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

- In a year after publication, 90% of researchers who were asked to provide their raw data responded.
- What was the response rate 10 years later?
- How many researchers responded at all?
- If you need raw data for a meta-analysis, how likely are you to find it?



of your

e?

they be



## 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.

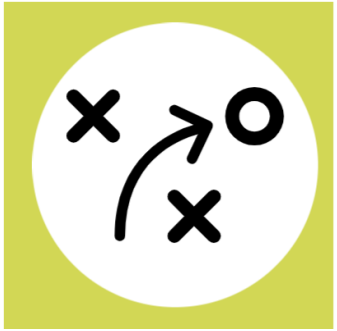
# What went wrong for Dave

- One of the external drives fails, leading to the loss of some of Dave's data prior to analysis being completed. This data loss is not discovered for several weeks. **There is no back up of this data.**
- This leaves Dave with two choices:
  - Recollect that data which may take a significant amount of time, or
  - Publish what he can, even though the power of the study has been reduced



# What could you do to be better?

<b>Make a plan for your data</b>	Prepare yourself for future challenges and problems
<b>Organize and document your data consistently</b>	Save yourself time and resources in the future.
<b>Store and back-up your data securely</b>	Avoid loss of data from theft, corruption, or failure of storage devices.
<b>Make sure your data is ready for archival and sharing</b>	increase the accessibility of your research and allow others to reproduce and use your results.



# 1. Make a plan for your data

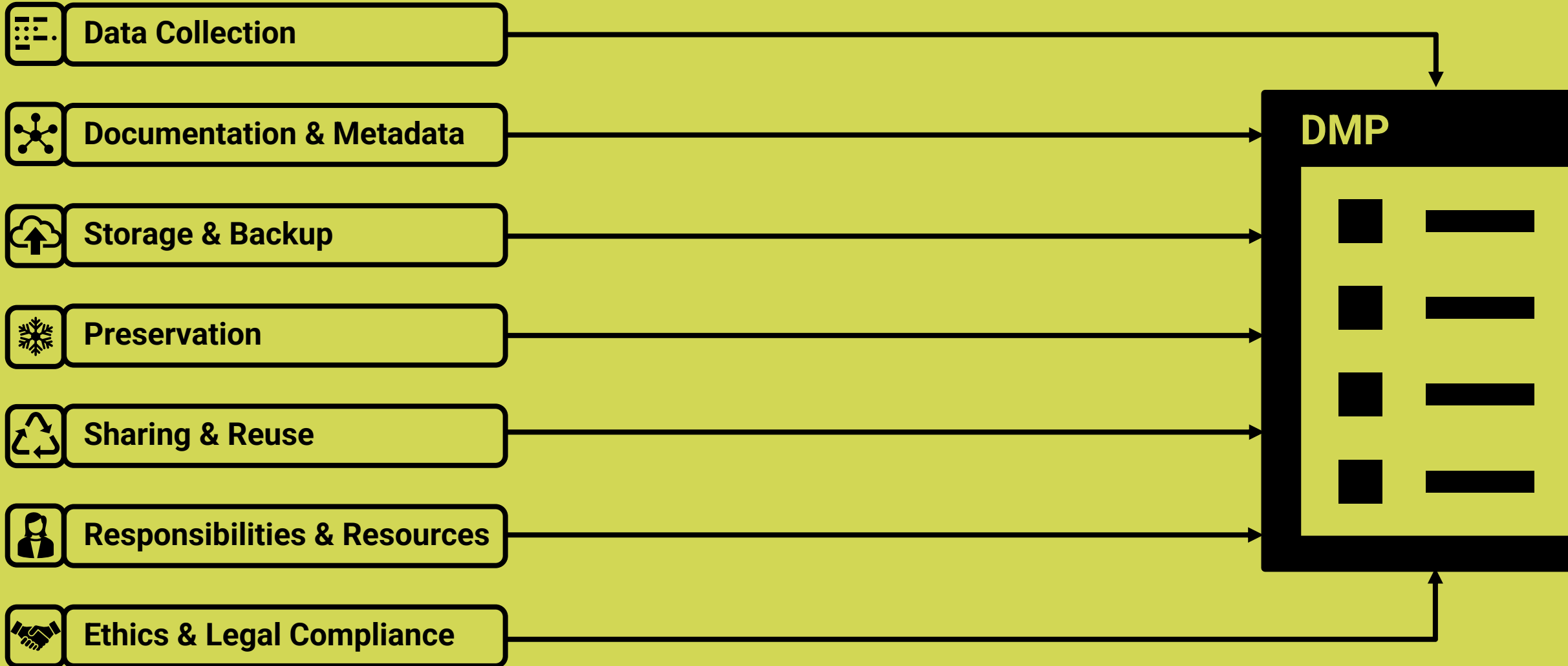
- A **Data Management Plan (DMP)** is your plan for how you will create, store, organize, document, secure, preserve, and share your research data.
- A document which speaks to the management of data both **during** the active phases of your research and **after** the completion of the research project.

# Why create a DMP?

- A **living document** – something you'll work with, adapt, and change through your research.
- Create it at the start of your research - avoid pitfalls and problems before they occur.
- Prepare for future stages of research including potential data sharing (if desired).
- Research is a team effort – collaborate on your DMP.
- Many research funders require grant applicants to submit a DMP – including the Tri-Agencies (NSERC, CIHR, SSHRC – started 2022), NIH, and others.

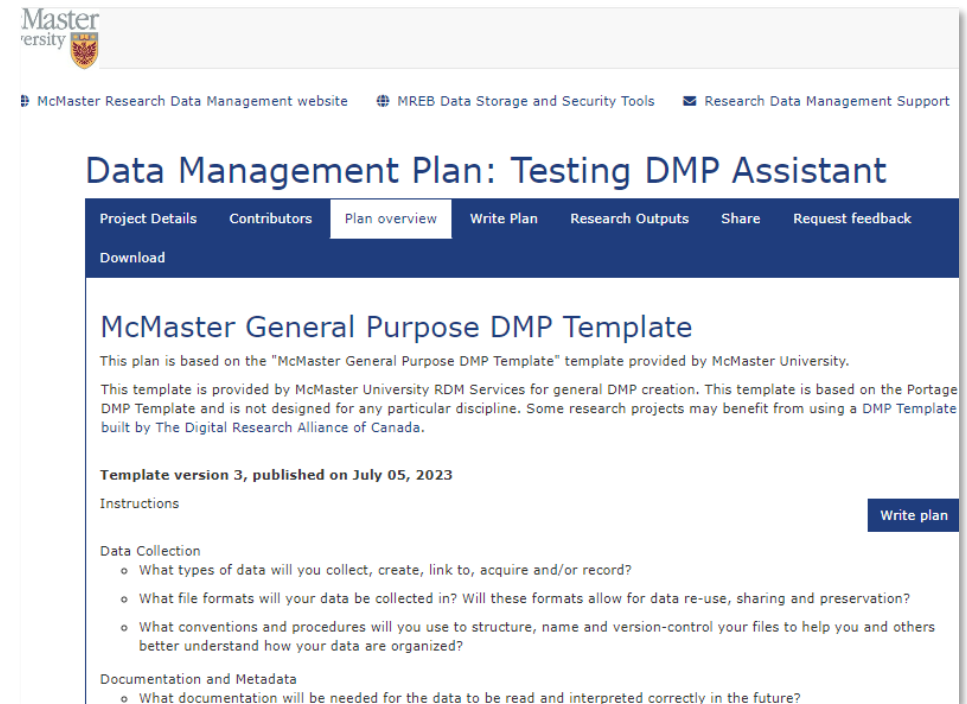


# What goes in a Data Management Plan?



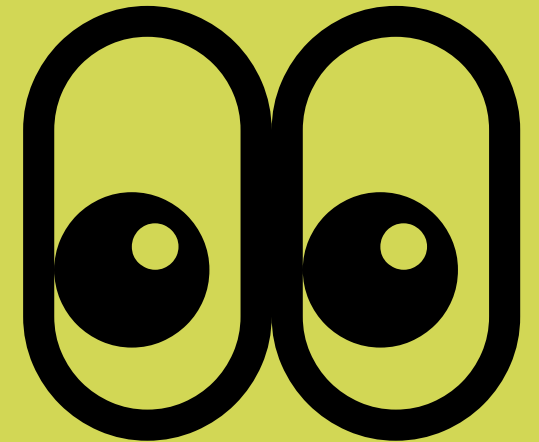
# DMP Assistant

- A web-based, bilingual data management planning tool.
- Available to all researchers in Canada.
- Walks you through relevant questions for data management.
- Templates and exportable plans.
- Send to RDM Services for review!
- Access at [rdm.mcmaster.ca/plan](https://rdm.mcmaster.ca/plan)





**Who do you think could  
potentially be involved in  
reviewing a DMP?**

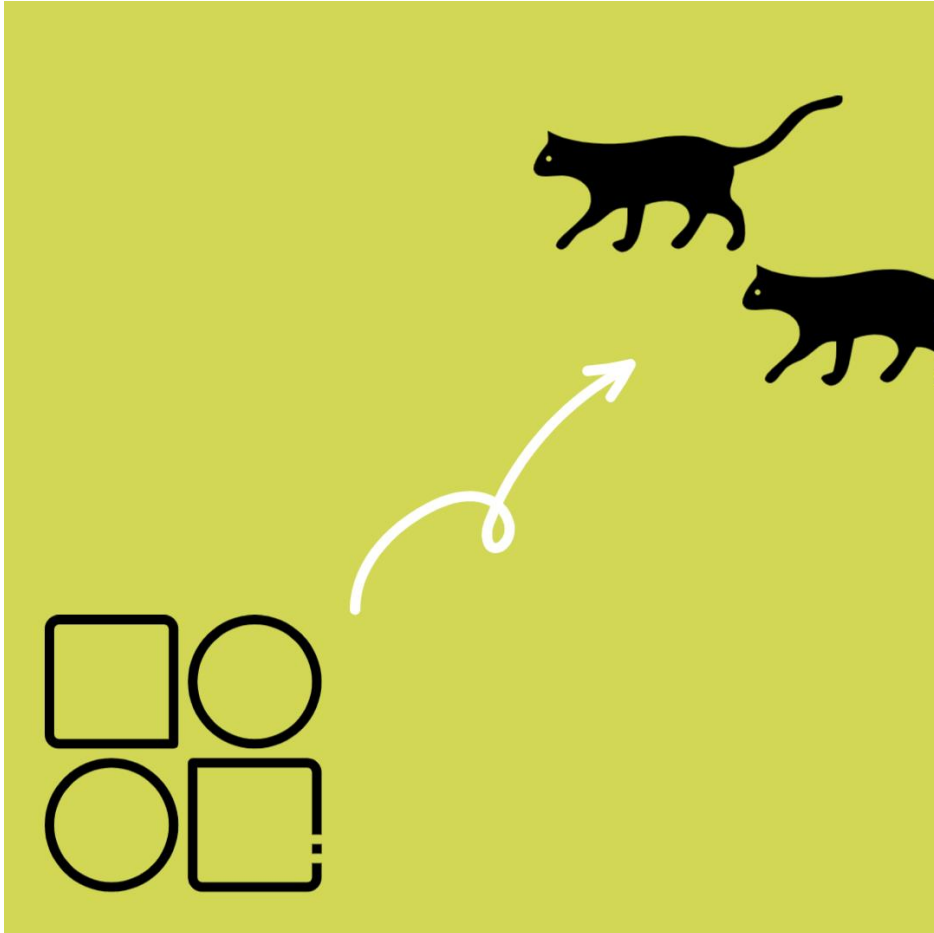




## 2. Organize and document your data consistently

- 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?

# What were you thinking?



*What did you mean when you made this variable “cat2”? Did it mean cat? Category? Something else?).*

Raw data can be complicated. Numbers in a spreadsheet can be hard to interpret, variable names might have abbreviations that make it hard for others to understand what it truly is.

Documentation and organization make data easier to understand and reuse.

Example from Samantha Ming, “Bad Variable Names to Avoid,” Medium, October 8 2018. <https://medium.com/@samanthaming/bad-variable-names-to-avoid-%EF%B8%8F-bbab6614f36>

## Soil radiometrics: Field and remote data sets for model building and validation

242.3KB

Contributors: [Cassia Read](#), [David H. Duncan](#), [Chiu Yee Catherine Ho](#), [Matt D. White](#), [Peter A. Vesk](#)

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 et al. (2017) Useful surrogates of soil texture for plant ecologists from airborne gamma-ray data. 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. (2017) *Useful surrogates of soil texture for plant ecologists from airborne gamma-ray data*. See the respective data set wiki pages for further information on provenance and metadata.

### Files

Name	Modified
Soil radiometrics: Field and remotely-sensed data sets for model building and validation	
OSF Storage (United States)	
Location.png	2017-05-

Cassia Read, David H Duncan, Chiu Y C Ho, Matt D White, and Peter A Vesk, "Soil Radiometrics: Field and Remotely-Sensed Data Sets for Model Building and Validation," OSF, June 15, 2018, [osf.io/uaç6x](https://osf.io/uaç6x).

# 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](https://rdm.mcmaster.ca/organize)

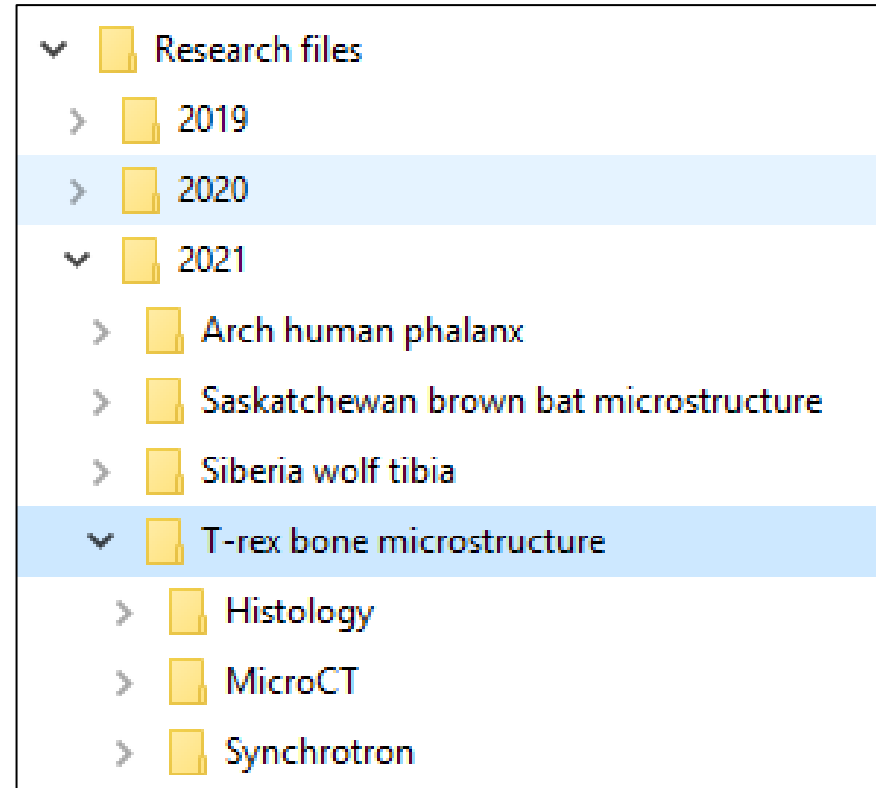
# Keeping folders organized makes it easier to find things.

Make organizing files and folders a habit – that way, it's easy to know where things go.

File organization schemes can include:

- By project
- By researcher
- By experiment type
- By date (often year)
- By some combination of the above

*(i.e. 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).
- *e.g. 2023\_02\_08\_YouthOutreach\_GuelphStaff\_TM.wav*
- **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.

*tip: if you start your file name with dates like 2023-10-03-filename they will sort by date automatically*

**Date:** 2023\_02\_08 (collection date)  
**Project Name:** YouthOutreach  
**Short Description:** GuelphStaff  
**Name:** TM (Tracy MacDern)

**What issues can we see with these files? What did this person do well? What could they have done better?**



A STORY TOLD IN FILE NAMES:

Location: C:\user\research\data

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

Type: Ph.D Thesis Modified: too many times Copyright: Jorge Cham www.phdcomics.com

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

Kristin Briney, "README.TXT," Data Ab Initio, February 25, 2014 <http://dataabinitio.com/?p=378>

## 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.



## 4. Store and back-up your data securely

- **Data Loss**
  - Theft of devices
  - Loss of devices
  - Accidental damage or destruction
  - A USB drive is not an archive!
- **IT Security:**
  - Computer viruses, malware, ransomware.

# Even cautious people can lose data.



*Dr. X leads a large, well-funded lab that conducts multiple expansive research projects. Because the lab is so busy, Dr. X and their team also receive many emails every day. Although they have taken some IT training, one of them clicks a link in a carefully-worded email and their research system becomes infected. All of Dr. X's critical research files are encrypted; locked down until the university pays a ransom. Depending on how well Dr. X's research is backed up, they could be facing an impossible decision: lose years of research...or pay up.*

# Avoiding Data Loss: **3-2-1 Backup Strategy**

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

<https://rdm.mcmaster.ca/finder>

- If you're looking for a good place to store research data, try our **Research Data Storage Finder** webtool, where you can get personalized recommendations.
- Answer questions to narrow down options depending on risk, volume, and other needs.
- This tool also allows you to compare feature sets of McMaster-approved options.
- Webinar on data storage:

<https://scds.github.io/intro-rdm/storage.html>

The screenshot shows the 'Data Storage Finder' web application. It includes a title, an introductory paragraph about the tool's testing phase and purpose, and three numbered steps: 1. Answering questions about storage needs, 2. Selecting providers to compare, and 3. Exploring provider details. The interface is divided into two main sections. The left section, 'Step 1: Answer these questions to narrow down storage provider options.', contains a 'Clear Answers' button and two questions: '1. What risk level is your data?' with radio buttons for Low, Medium, and High, and '2. What type of data storage are you looking for?'. The right section, 'Step 2: Select data storage providers you would like to compare', features a 'Select All' button and a grid of provider cards. Each card has a title, a brief description, and a selection radio button. The visible cards include: Compute Canada (Advanced research computing systems, storage and software), Compute Canada NextCloud (Advanced research computing File hosting services), Datarverse (Store, share, publish and discover research data), FRDR (Find and Share Research Data), Github (Distributed version control system for software code), MacDrive (File Synchronization and Sharing solution), MacDrive (Encrypted Data) (Encrypted File Synchronization and Sharing solution), and MacDrop (Web service transfer file).

## **Encrypt sensitive data:**

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

**Encryption** is a process of transforming information so that it is only readable to a person with the correct authorization.

- 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.

For more details see our webpage: <https://rdm.mcmaster.ca/secure>

# 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
- Most 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!
- ✓ **Devices:** Use a strong password on your computer and phone, too
- ✓ **Monitored:** Watch for data breaches in platforms you use and change your password when they happen.



**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	<a href="mailto:example@mcmaster.ca">example@mcmaster.ca</a>	football
DropBox	<a href="mailto:example@mcmaster.ca">example@mcmaster.ca</a>	football
Google	<a href="mailto:test.example@gmail.com">test.example@gmail.com</a>	football
Instagram	<a href="mailto:test.example@gmail.com">test.example@gmail.com</a>	football1
ORCID	<a href="mailto:example@mcmaster.ca">example@mcmaster.ca</a>	football
Zotero	<a href="mailto:example@mcmaster.ca">example@mcmaster.ca</a>	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?

**What is the worst password you can imagine?** *(add plus signs to the ones you think are the worst! winner wins a tiny croc to put on your crocs!)*





## 5. Make sure your data is ready for archival and sharing

- **Data Sharing:** Open and free data sharing supports research ideals like transparency, accessibility, reproducibility, collaboration, and maximizes the impact and visibility of research.
- **Data Archiving:** 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?



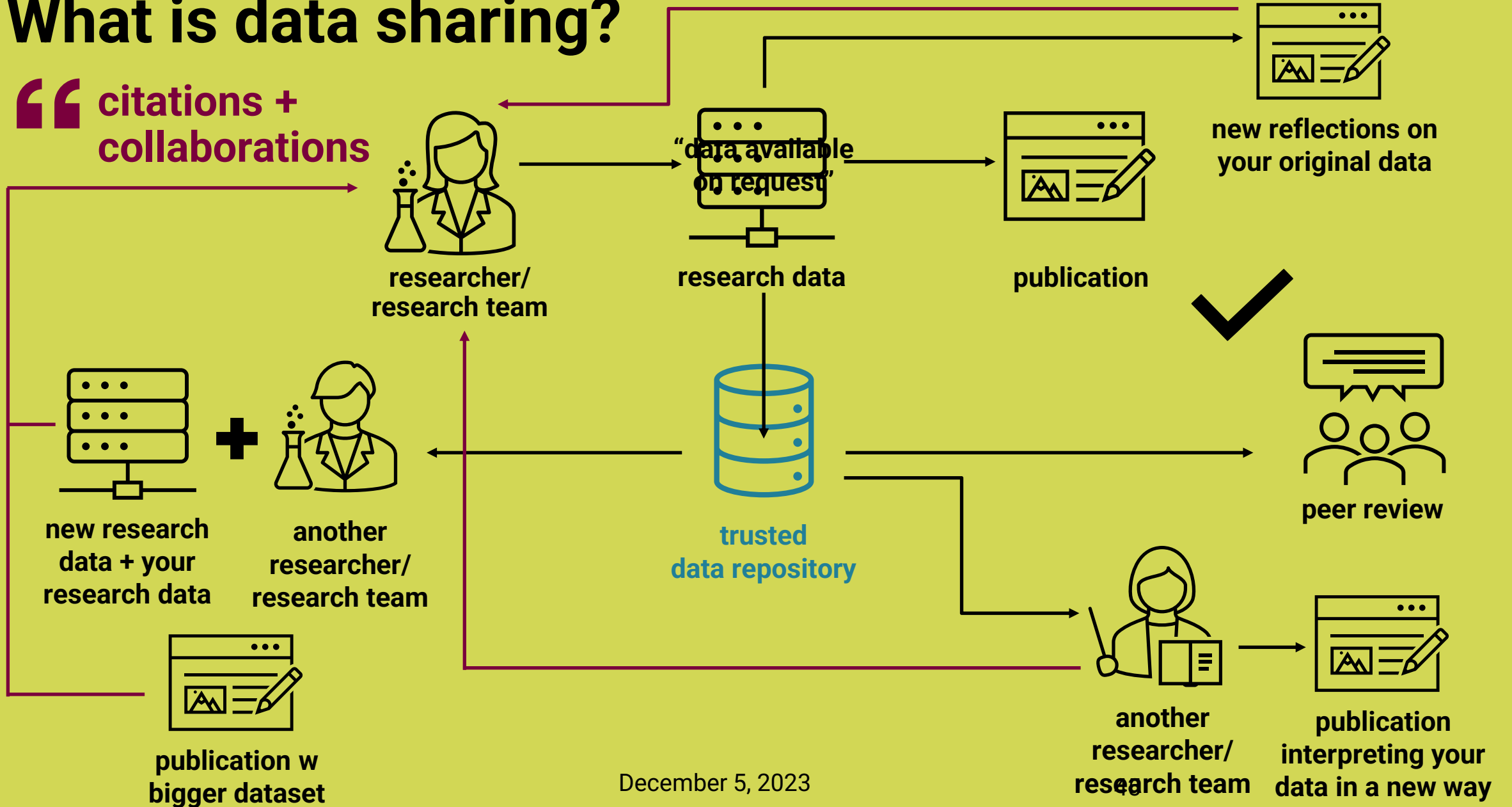
# Open Research



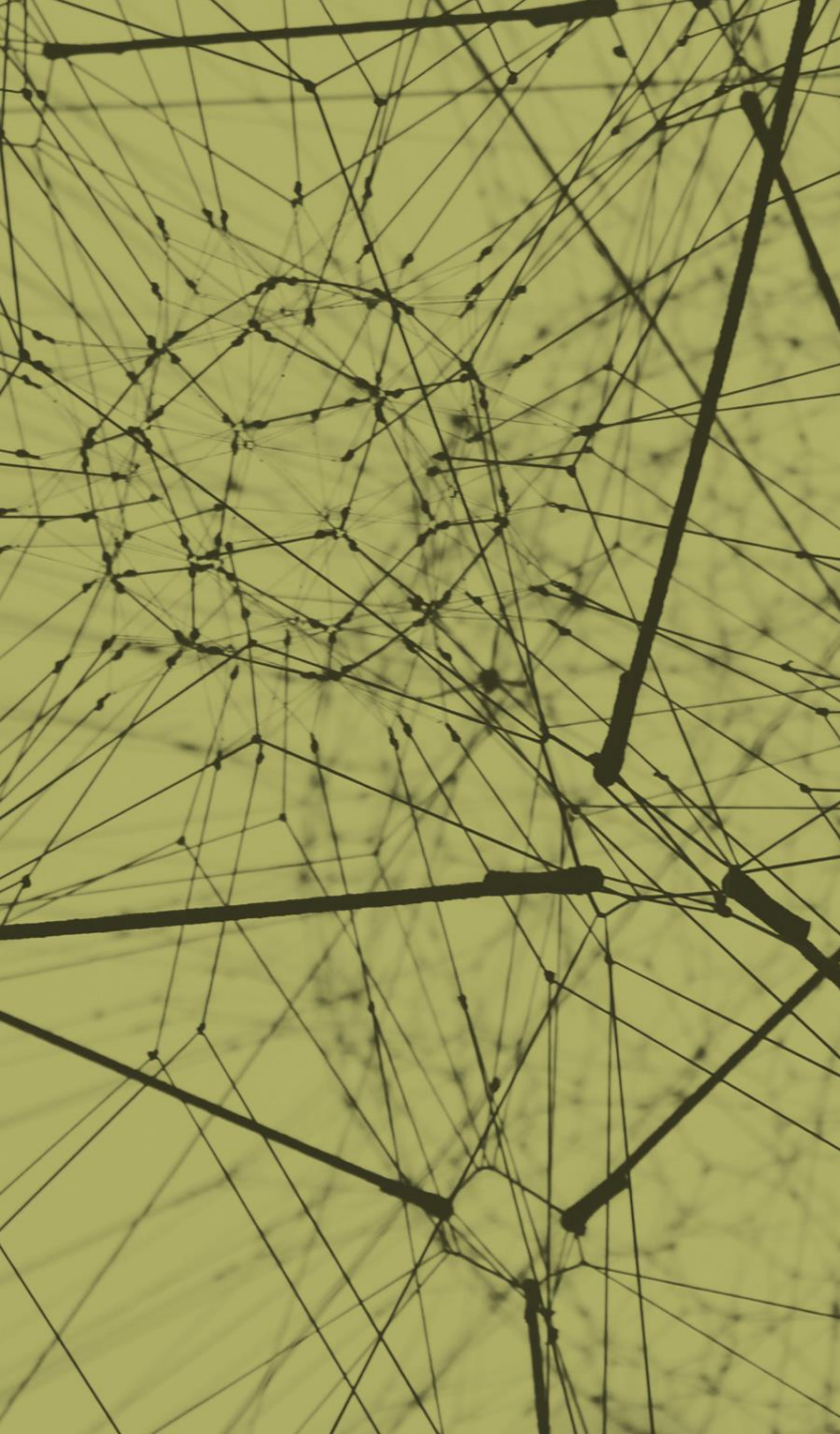
- Supports **reproducibility** and **replicability** – heightens confidence in results and supports research integrity
- Leads to **new collaborations** – potential for meta-analyses over a wider topic area.
- Research is often publicly funded – data as a “public good” that is accessible by journalists, other researchers, etc.
- Limits **over-researching communities** for qualitative studies. Share datasets with **community partners** – value added, support long-lasting research relationships.

# What is data sharing?

“ citations + collaborations ”



December 5, 2023

A complex network graph with numerous nodes and edges, rendered in a dark green color against a lighter green background. The graph is dense and interconnected, with some nodes having significantly more connections than others.

# 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 Funder 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? Funder Requirements

The screenshot shows the Government of Canada website. The header includes the Canadian flag, the text "Government of Canada / Gouvernement du Canada", and a search bar. A navigation menu is visible on the left. The main content area is titled "Tri-Agency Research Data Management Policy" and includes a "1. Preamble" section. The preamble text states: "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 professor and disciplinary standards, domestically and internationally. These standards support research excellence by ensuring that research is performed ethically and makes good 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 screenshot shows the NIH Scientific Data Sharing website. The header includes the NIH logo and the text "SCIENTIFIC DATA SHARING". A search bar is visible. A navigation menu on the left lists "DATA MANAGEMENT AND SHARING POLICY", "GENOMIC DATA SHARING POLICY", "OTHER SHARING POLICIES", "ACCESSING DATA", and "ABOUT". The main content area is titled "Data Management and Sharing Policy" and includes a paragraph: "NIH has a longstanding commitment to making the results of NIH-funded research available. Responsible data management and sharing has many benefits, including accelerating the pace of biomedical research, enabling validation of research results, and providing accessibility to high-value datasets." A link "About the Data Management and Sharing Policy" is visible at the bottom.

## ■ Tri-Agency RDM Policy – CIHR:

- **Data Management Plans:** Network Grants in Skin Health and Muscular Dystrophy, Virtual Care/Digital Health Team Grants, Data Science for Equity
- **Data Deposit** - CIHR currently requires researchers to “deposit bioinformatics, atomic, and molecular coordinate data into the appropriate public database.”
- See the [Tri-Agency Data Management Policy](#) for details.

## ■ National Institutes of Health (NIH)

- **Data Management Plans:** “NIH requires all applicants planning to generate scientific data to prepare a DMS Plan that describes how the scientific data will be managed and shared.”
- **Data Sharing:** “Scientific data should be made accessible as soon as possible.”
- See the [NIH Data Management and Sharing Policy](#) for details.



# Data Sharing – Archival and Preservation

Sharing data in a research data repository is a good way to ensure that data will be accessible over the long term.

- Datasets are given **DOIs**.
- Datasets are in platforms that have plans for long term storage.
- Datasets are published with enough documentation to enable re-use or re-analysis.

# Data Repository Decision-Making

Publishing data in a recognized data repository is the best way to share data. There are thousands of data repositories.



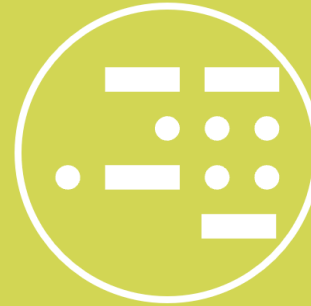
## Domain Specific Repositories

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>

# 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 website. At the top, there is a navigation bar with the Borealis logo, search, user guide, support, language, and login options. The main header features the McMaster University logo and navigation links for the Dataverse and RDM Services. Below this, there is an 'About Dataverse' section with introductory text and links to research data management resources. A search bar is present with a search button and an 'Add Data' button. On the left, there are filters for 'Dataverses (19)', 'Datasets (64)', and 'Files (783)', along with a 'Dataverse Category' list. The main content area displays search results, with the first result being 'Long-term Rocky Tidal Community Data from Discovery Bay, Jamaica' by Jurek, Kolasa, dated June 30, 2022. A second result, 'Nurr1 is not an essential regulator of BDNF in mouse cortical neurons', is partially visible below.

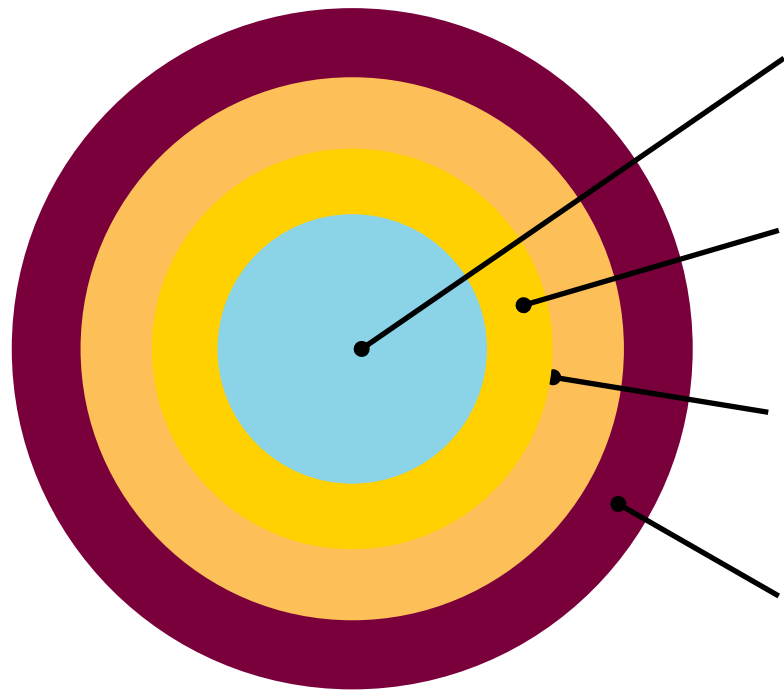
# Federated Research Data Repository (FRDR)

<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



# 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

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

Persistent Identifiers (PIDs) 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.



# Standards - File Formats

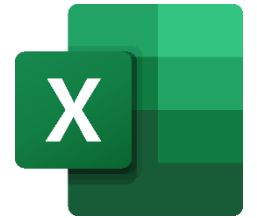
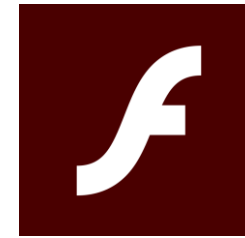
*Have you ever saved data on a CD, DVD, or BluRay? How about a zip disk or HD DVD? Do you use Google Docs or Prezi where all your documents are stored online on a proprietary platform in a proprietary format? What would you do if that platform closed down?*

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

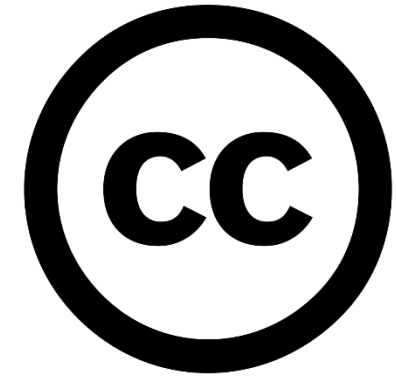
[https://site.uit.no/dataverseno/deposit/prepare/  
#what-are-preferred-file-formats](https://site.uit.no/dataverseno/deposit/prepare/#what-are-preferred-file-formats)



# Metadata - Data Licensing

- **Creative Commons (CC)** - ([creativecommons.org](https://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](https://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](https://localcontexts.org/licenses)

Images from Sq'ewlets, "Traditional Knowledge Labels," [http://digitalsqewlets.ca/traditional-knowledge\\_connaissances\\_traditionnelles-eng.php](http://digitalsqewlets.ca/traditional-knowledge_connaissances_traditionnelles-eng.php) and [Creative Commons](https://creativecommons.org/licenses/by/4.0/), fixed by [Quibik](https://www.quibik.com/).



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:xa in our language. This label indicates that there is additional knowledge about a certain subject that cannot be shared on the website.



# Metadata - Tags + Documentation

Metadata is “data about data” - keywords and tags to help with searching, or documents that explain your decision making and organization.

**Metadata:** When you deposit data in a repository, include metadata such as a title, date the data was collected, language, description, keywords, geographic location, and more. This helps other researchers narrow down the search and understand if the data is useful to them.

**Documentation:** Deposit a README file alongside your dataset to ensure others can understand how you gathered the data and how they can use it.

<https://rdm.mcmaster.ca/publish>

This checklist augments our more comprehensive [McMaster Dataverse Data Deposit Guidelines](#) and [McMaster Dataverse Deposit Walkthrough](#) documents. *Items marked with an asterisk \* are mandatory for submissions.*

## Before Deposit

- \*README:** Develop a README file to submit alongside dataset [template, [more info](#)]. You do not need to use the template but information equivalent to sections 1-3 from the template is required.
- Documentation:** Gather all the documentation for your dataset, potentially including: data dictionary, data model, code book, interview guide, etc. [\[more info\]](#)
- \*Anonymization/De-identification:** McMaster Dataverse does NOT accept datasets containing confidential or sensitive information. Remove, replace, or redact data until they are de-identified and non-confidential. [\[more info\]](#)
- \*Ethics Approval:** Confirm you have MREB or HiREB approval to share data (*if applicable*).
- \*Citation and Credits:** Confirm you have credited, linked, and cited third-party sources, including data, code, or software (*if applicable*)
- Dataset organization:** Use consistent file naming and folder organization. A well-structured dataset is easier to understand and share. [\[more info\]](#)
- Sustainable file formats:** Make your data files accessible long-term and for re-use. [\[more info\]](#)

## During Deposit

- \*Account + New Dataset:** [Access McMaster Dataverse](#) and set up your account by clicking “Log In” at the top of the page. Click “Add Data” and “New Dataset”.
- \*Data License:** Select an open license under “Dataset Template.” [\[more info\]](#)
- \*Add Metadata:** Ensure your data is findable by adding descriptive information.
  - \*Basic Metadata:** Add Title, Author, Contact, Description, and Subject. [\[more info\]](#)
  - \*Publication:** Enter citation information to Related Publication(s) – *min. title + journal*.
  - Time:** Add details to “Time Period Covered” and “Date of Collection” (*if applicable*)
  - Place:** Add details to “Geospatial Metadata” (*if applicable*)
- Researcher ID:** Connect your research by adding an ORCID or another researcher ID for yourself and your collaborators. [\[more info\]](#)
- \*Dataset:** Upload data files under the “Files” heading – click “+ Select Files to Add” – *.zip files are unpacked automatically. If you have a lot of files, compress the whole directory in a .zip and upload that file to maintain file/folder structure.*
- File Names + Tags:** Modify file names and add text descriptions. *If desired, add tags for clarity –*

# Open Access (OA) Publishing

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

## Publication 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, but there are lots of resources available. Learn more about Open Access here:  
<https://library.mcmaster.ca/research/open-access#tab-open-access-publishing>

## Journals

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

## MacSphere

- McMaster's institutional publication repository. This is where graduates publish theses!
- Publish your data in McMaster Dataverse, publication in MacSphere, link both. We can help!



**What are some opportunities you can think about with sharing data? What are some hesitations you might have?**



# Top 4 ideas for improving your research data management:



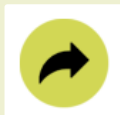
Make a **plan** for data management.



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

- Teams channel hosting asynchronous discussion
- Monthly meetings featuring presentations from researchers across the university on how they do data management
- **Thursday, September 28 – 11 AM -12 PM. DMP Roundtable:** Natalie Harvey-Younis (Psychiatry & Behavioural Neurosciences); Benjamin Taylor (MacPherson Institute); Melissa Parker and Saleha Ansari (Rehabilitation Science)



All are welcome to join! <https://u.mcmaster.ca/rdm-community>

October 25, 2023 | 10:30am-12pm  
Virtual Workshop + Sandbox Session

# Data Management Plans & Intro to DMP Assistant

[u.mcmaster.ca/scds-events](https://u.mcmaster.ca/scds-events)



December 5, 2023

**SCDS**

62

Library



November 29, 2023 | 10:30am-12pm  
Virtual Workshop + Sandbox Session

# Depositing & Sharing Data Online with McMaster Dataverse

[u.mcmaster.ca/scds-events](https://u.mcmaster.ca/scds-events)



December 5, 2023

**SCDS**

63

Library

