# Best practices for managing data in your research

Isaac Pratt, PhD + Danica Evering, MA

October 26, 2022





**Research Data Management** 



#### **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: <u>scds.ca/events/code-of-conduct/</u>





2

### **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.





3

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

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





4



McMaster University sits on the traditional territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the "Dish With One Spoon" wampum agreem<u>ent.</u>

#### What is Research Data Management anyways?

scds.ca



Library

#### 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?

an Centre for Digital Scholarship

scds.ca





#### FOR MY LOST LAPTOP

abar all the details of your

1 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. and phone number is

PLEASE contact me and I would appreciate it so so much!!!

ble

Library

#### Is your data vulnerable?





#### Is your data vulnera

University of Manitoba Psychology



## In a year or two, wi experiments? Have

- What will happen t
- How much work w
- If you needed to show to understand it with



the details of your ails?

e/move/retire?

your data?

Winnipeg Free Press

tor, would they be able



Library



#### Lets 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



Librarv



#### 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 (starting later this year)





#### What goes in a data management plan?

scds.ca



Library



#### Alliance de recherche numérique du Canada

Welcome to DMP Assistant.	Sign in Create account
leveloped by the Digital Research Alliance of Canada (the Alliance) in Ollaboration with host institution University of Alberta to assist researchers in reparing data management plans (DMPs). This too is freely available to all esearchers, and develops a DMP through a series of key data management questions, upported by best-practice guidance and examples.	* Password
MPs are one of the foundations of good research data management (RDM), an nternational best practice, and increasingly required by institutions and funders, ncluding the Canadian Tri-Agencies as outlined in their Research Data Management olicy.	Forgot password?
Setting started: brief Guide – Data Management Plans brief Guide – Create an Effective Data Management Plan Primer – Data Management Plan Horo to Manage Your Data introduction to Data Management Plans (DMPs) brind to DMP Assistant Managing DMPs with DMP Assistant Webinars: support Your Research with DMP Assistant 2.0 support Your Research with DAta Management Planning	Jagu W
for more resources and training materials spanning the entire research data life cycle, ee the Portage Network Training Resources.	
ool, and uses the DMP Roadmap codebase developed by DCC and the University of	
Digital Research   Alliance de recherche GMPONLINE CARL @ UNIV Alliance of Canada   numérique du Canada	ARY OF ALBERTA *UC3

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 analyse 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?



Sherman Centre scds.ca ≡

Modifie

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

Description: Repository for model training and testing data sets for the article: Read Ci White M, Vesk 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

scds.ca

This project is home to the soil data for north-west Victoria, Australia used by Read e Useful surrogates of soil texture for plant ecologists from airborne gamma-ray dete. See the respective data set wiki pages for further information on provenance and

Files	
Name 🔨 🗸	

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/uac6x.<sub>Data.xisx</sub> Sherman Lentre for Digital Scholarship

#### **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





Do you have files named like this?

Is this a good file name system?



A STORY TOLD IN FILE NAMES:			
Location: 😂 C:\user\research\data			~
Filename 🔺	Date Modified	Size	Туре
<ul> <li>data_2010.05.28_test.dat</li> <li>data_2010.05.28_re-test.dat</li> <li>data_2010.05.28_re-re-test.dat</li> <li>data_2010.05.28_calibrate.dat</li> <li>data_2010.05.28_huh??.dat</li> <li>data_2010.05.28_WTF.dat</li> <li>data_2010.05.29_aaarrrdb.dat</li> <li>data_2010.05.29_dat</li> <li>data_2010.05.29_orap.dat</li> <li>data_2010.05.29_notbad.dat</li> <li>data_2010.05.29_Woohoo!!.dat</li> <li>data_2010.05.29_USETHISONE.dat</li> <li>analysis_graphs.xls</li> <li>ThesisOutline!.doc</li> <li>Notes_Meeting_with_ProfSmth.txt</li> <li>JONK.</li> </ul>	3:37 PM 5/28/2010 4:29 PM 5/28/2010 5:43 PM 5/28/2010 7:17 PM 5/28/2010 9:58 PM 5/28/2010 9:58 PM 5/28/2010 12:37 AM 5/29/2010 2:40 AM 5/29/2010 3:22 AM 5/29/2010 4:16 AM 5/29/2010 4:47 AM 5/29/2010 5:08 AM 5/29/2010 7:13 AM 5/29/2010 7:26 AM 5/29/2010 11:38 AM 5/29/2010 2:45 PM 5/29/2010	420 KB 421 KB 420 KB 1,256 KB 30 KB 30 KB 30 KB 437 KB 670 KB 1,349 KB 2,894 KB 455 KB 38 KB 1,673 KB	DAT file DAT file TAT file DAT file
😝 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.phd	comics.com



∟ibrary

Project: Kristin's important chemistry project Date: June 2013-April 2014 Description: Description of mu 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, <sup>=</sup>ebruary 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 notetaking software, reference management software, and collaboration software.

#### Storing your data safely

Avoiding data loss

- RIP)
- Theft or loss of devices, accidental damage or destruction
- 2. IT Security
  - 。 Computer viruses, malware, ransomware



1.



### **Backup Strategies (3-2-1)**

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



scds.ca

Copies of your data (at least!)

#### Example:

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



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

- a "production" (working) copy
- a "production backup" copy



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.







## 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 <u>VeraCrypt</u>.

#### 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

#### 13.25

Microsoft

## Approve sign-in

Open your Microsoft / approve the request to

an't use my Microsoft Auther





#### **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 LastPass.







Here are some common mistakes people make when creating passwords:

🚺 Au	toSave 🔵 Off) 📙 Test E	xample's Passwords.xlsx • Saved 🗸
File	Home Insert Draw	Page Layout Formulas Data
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

Have you made any of these mistakes before?

**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.

**Too Simple:** Simple passwords composed of common words are easy to guess. *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.





#### **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

#### Why should I share my data?

Improve the quality of your research

- Allow verification of results/code by peers
- Potential of 'mega' datasets

Improve the value of your research

- Avoid duplication of data collection or programming
- Maximizes use of your data/code

Improve the impact of your work

- Increases the visibility of research
- Can lead to new collaborations and partnerships









## 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 <u>https://doi.org/10.1371/journal.pone.0230416</u>
- 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 <u>https://doi.org/10.7717/peerj.175</u>







# Why share data? Journal and Publisher Requirements

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

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

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





## Why share data? Tri-Agency Requirements

Governme of Canada

Government Gouvernement of Canada du Canada

ient

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 <u>Canadian Institutes of Health Research (CIHR)</u>, the <u>Natural Sciences and Engineer</u> <u>Research Council of Canada (NSERC)</u>, and the <u>Social Sciences and Humanities Researc</u> <u>Council of Canada (SSHRC)</u> (the agencies) are federal granting agencies that promote support research, research training, knowledge transfer and innovation within Canad.

The agencies expect the research they fund to be conducted to the highest professior 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 resea excellence.

The agencies believe that research data collected through the use of public funds sho 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 <u>Tri-Agency Data Management Policy</u> 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 <u>arXiv.org</u>)
- Researcher is responsible to navigate copyright requirements of the journal

<u>Journals</u>

- 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 (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.

Connecting research and researchers



Photo by Nasa on Unsplash. MCMaster University

Library



#### **Publishing Data: File formats**

Data should be stored in sustainable file formats and media

- Have you ever saved data on a CD, DVD, or BluRay? How about a zip disk or HD DVD?
- Do you use an online document processing software like 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?
- $_{\circ}$   $\,$  Adobe Flash was shut down December 31st 2020  $\,$





## 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





#### **Sustainable File Formats**

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:

- $_{\circ}$  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 <u>https://site.uit.no/dataverseno/deposit/prepare/</u> <u>#what-are-preferred-file-formats</u>







erman Centre for Digital Scholarship

Lewis & Ruth

scds.ca

Library

#### 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'éwlets, "Traditional Knowledge Labels," <u>http://digitalsqewlets.ca/traditional-knowledge\_connaissances\_traditionnelles-eng.php</u> and <u>Creative Commons</u>, fixed by <u>Quibik</u>.







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.



#### **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/ nmunity Meeting 2022 1 June 14, 15, and 16

Speakers and Chairs Registration





#### #Dataverse2022

Healy's professional background is multis worked in numerous health capacities at , and international levels. Actively involved ways of knowing, Bonnie's passion is to itiles and provide them with tools that they port communities in information data esearch methodologies. Bonnie's irst Nations information systems gives her ong 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.

**Datavers** 

The Dataverse Community Meeting is hosted by Harvard's Institute for <u>Quantitative Social Science</u>. Learn more about The Dataverse Project at our <u>dataverse.org site</u>.







### 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 designed for researchers.





## **Data Repository Decision-Making**

Publishing data in a recognized data repository is the best way to share data. There are thousands of data repositories. <u>MacSphere is our institutional repository for publications</u>.



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 codespecific 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://repositoryfind er.datacite.org



Library



## 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

堡 borealis	Search 👻 U	Jser Guide Support English - Log
	McMaster University	
McMaster University Datav	erse (McMaster University) McMaster RDM Services	
Borealis >		
		🗷 Contact 🛛 🔁 Sha
About Dataverse	3	
The McMaster University Dataverse choose to make content available p	a is a research data repository for our faculty, students, and staff. Files are held in a secure environmenublicly, to specific individuals, or to keep it private.	ent on Canadian servers. Researchers can
For information on research data m	anagement, including best practices for sharing and preserving research data, consult McMaster's R	esearch Data Management website.
The Library can assist with using th	e repository service and will help you prepare your data for sharing and preservation.	
<ul> <li>Before depositing a datase</li> <li>For a full walkthrough of department</li> </ul>	t in the McMaster Dataverse, please look through the Data Deposit Guidelines for information epositing data to McMaster Dataverse, please click here	on what is required to deposit data.
If you need assistance please conta	act: rdm@mcmaster.ca	
Search this dataverse	O Advanced Search	+ Add Data
ocarcin inio dataverse	Advanced Search	
✓ 8 Dataverses (19)	1 to 10 of 83 Results	Lt Sort
	I to 10 of 83 Results      Long-term Rocky Tidal Community Data from Discovery Bay, Jamaica     Jun 30, 2022     Junes, Kolasa, 2022, "Long-term Rocky Tidal Community Data from Discovery Bay, Jamaica Borealia. V1. UNF & Cca.2022G108/20VL oKFCGg== [lieUNF]      The rocky supertidal zone is home to a vide diversity of life. Understanding how individual speci- community and how the abidic environment can share these interactions cancels detailed study	It Sort-





Library

#### Ok so where do I put everything? FRDR

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

scds.ca

- 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 me: <u>rdm@mcmaster.ca</u>



## 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







# Thank You.

#### For more information:

Visit: rdm.mcmaster.ca

Contact me at: rdm@mcmaster.ca



**Research Data Management** 

McMaster



