



Research Data Management Workshop for HRS, MILO, ROADS

Isaac Pratt, PhD
June 11th, 2021



- McMaster University sits on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by

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 agreement.

HELLO! A bit about me:

- I am a Research Data Management Specialist working in the University Library and Research and High Performance Computing (RHPCS).
- My background is in Biological Anthropology, Medical Imaging, and Human Anatomy.
- I have a PhD in Anatomy & Cell Biology from the University of Saskatchewan.
- Email me with any RDM question or to set up a consultation:
pratti@mcmaster.ca

MY SERVICES

Consultations: We can meet with a researcher or research group to discuss RDM needs including:

- Data deposit for sharing and archival, with specific support for [McMaster Dataverse](#)
- Data Management Plan creation and management using the [Portage DMP Assistant Tool](#)
- Data storage & backup, data security, metadata & data documentation

Workshops: We can present targeted workshops to research groups, academic units, or other groups on any aspect of RDM.

Resources:

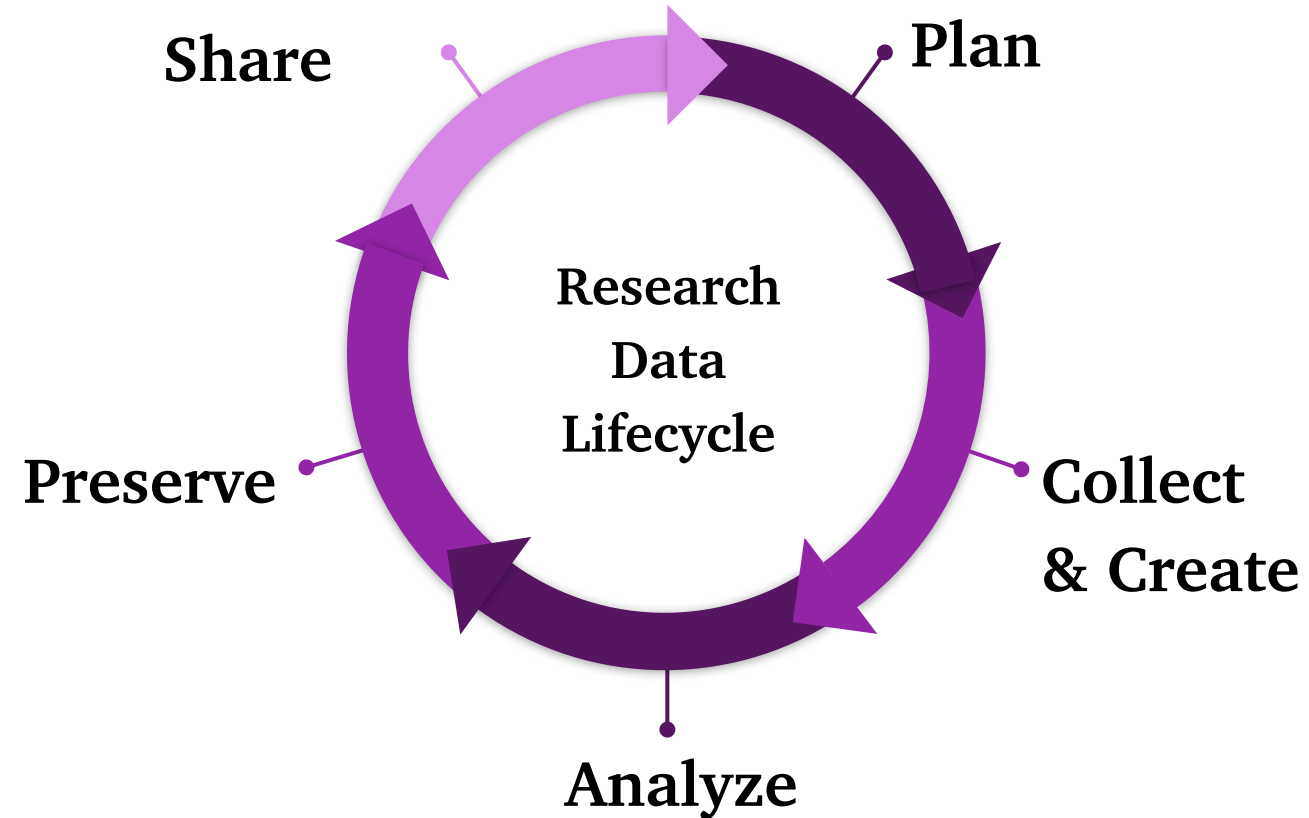
- RDM Website library.mcmaster.ca/services/rdm

OUTLINE FOR TODAY

- What is RDM?
- Tri-Agency RDM Policy
- Data Management Plans
- Data Management best practices
- Data Deposit

WHAT IS RESEARCH DATA MANAGEMENT ANYWAYS?

Research Data Management is the active organization & maintenance of data throughout the research data lifecycle to ensure its **security, accessibility, usability, and integrity**



WHY IS RESEARCH DATA MANAGEMENT IMPORTANT?

- Proper **data organization and planning** ahead saves time, resources.
- Good **data storage and backup** strategies help avoid loss of data from theft or aging/mechanical failure of storage devices.
- **Sharing data openly** allows others to reproduce and verify research results and increases the impact of research.
- **Depositing and Publishing data** increases the visibility of research

“BEST PRACTICES FOR MANAGING DATA IN YOUR RESEARCH” WEBINAR

Check out the recording of my webinar online at:
<https://scds.github.io/intro-rdm/intro.html>

Synchronous Workshop

Do **MORE**
with
Digital
Scholarship

**Best Practices for Managing
Data in your Research**

Lewis & Ruth
Sherman Centre
for Digital Scholarship

McMaster
University  Library



McMaster
University 

Research &
High Performance
Computing

McMaster
University 

Library



TRI-AGENCY POLICIES

Tri-Agency Statement of Principles on Digital Data Management: Expectations

- Data management planning
- Constraints and obligations
- Adherence to standards
- Collection and storage
- Metadata
- Preservation, retention and sharing
- Timeliness
- Acknowledgement and citation
- Efficient and cost-effective

Tri-Agency Statement of Principles on Digital Data Management <http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>

Tri-Agency Statement of Principles on Digital Data Management: **Responsibilities**

Researchers

- incorporating best practices
- developing DMPs
- adhering to policies and standards

Research Communities

- developing & promoting standards
- fostering excellence
- selecting repositories

Research Institutions

- supporting best practices
- providing access to resources
- creating guidance and policies

Research Funders

- developing policies & guidance
- promoting data management
- providing peer reviewers

Tri-Agency Statement of Principles on Digital Data Management: **Responsibilities**

Researchers

**To
Support**

- incorporating best practices
- developing DMPs
- adhering to policies and standards

Research Institutions

**To
Do**

- supporting best practices
- providing access to resources
- creating guidance and policies

Research Communities

- developing & promoting standards
- fostering excellence
- selecting repositories

Research Funders

- developing policies & guidance
- promoting data management
- providing peer reviewers



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 and support research, research training, knowledge transfer and innovation within Canada.

The agencies expect the research they fund to be conducted to the highest professional 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 obligations allow, available for reuse by others. To this end, the agencies support the FAIR

- “Research Data Management is a necessary part of research excellence.”
- “Recognize that data related to research by and with **First Nations, Métis, or Inuit** in Canada must be managed in accordance with data management principles developed and approved by these communities.”

**TRI-AGENCY
RDM POLICY**

**Institutional
Strategy**



DMPs



Deposit



Institutional Strategy



“Each postsecondary institution... is required to create an institutional RDM strategy and... make it publicly available on the institution’s website.”

“An institutional RDM strategy describes how the institution will provide its researchers with an environment that enables and supports RDM practices”

Research institutions must post their RDM strategies by **March 1, 2023.**

MCMASTER RDM INSTITUTIONAL STRATEGY WORKING GROUP

- The working group reports to the Research Information Technology committee.
- Group tasked with developing a draft Institutional RDM Strategy for early-mid 2022.
- Membership includes researchers from each Faculty, representatives from the ethics boards, libraries, and research support units (including ROADS/HRS/MILO).

DMPs



“For certain funding opportunities, the agencies will require data management plans (DMPs) to be submitted”

We provide guidance and templates for researchers using the **Portage DMP Assistant**

Initial set of grants due **Spring 2022**. Some pilot projects earlier.

Deposit

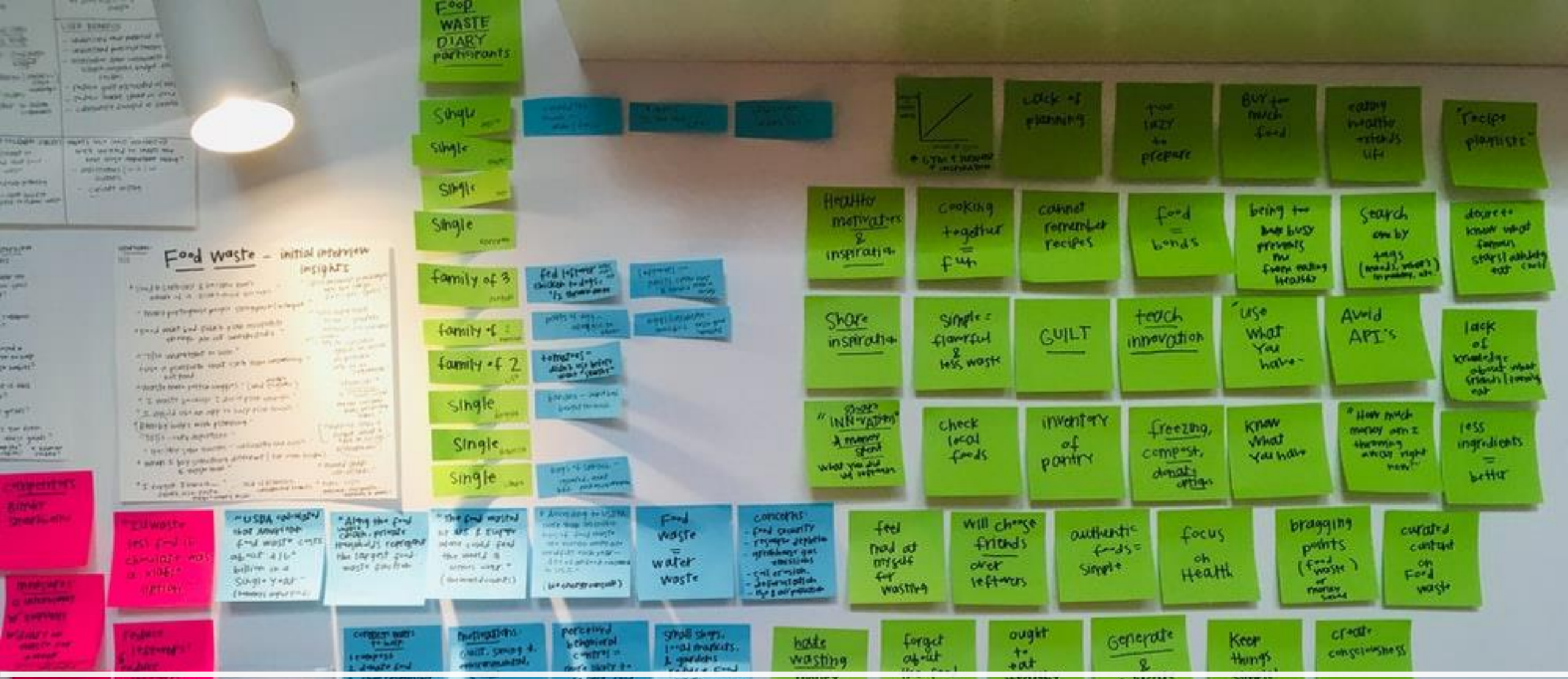


“Grant recipients are required to deposit into a digital repository all digital research data, metadata and code that support... journal publications and pre-prints”

We run a data repository for McMaster researchers:
the McMaster Dataverse

Note: Grant recipients will **not** be required to openly share their data

Dates still TBD. SSHRC and some CIHR data is already required to be deposited.



DATA MANAGEMENT PLANS



DATA MANAGEMENT PLANNING

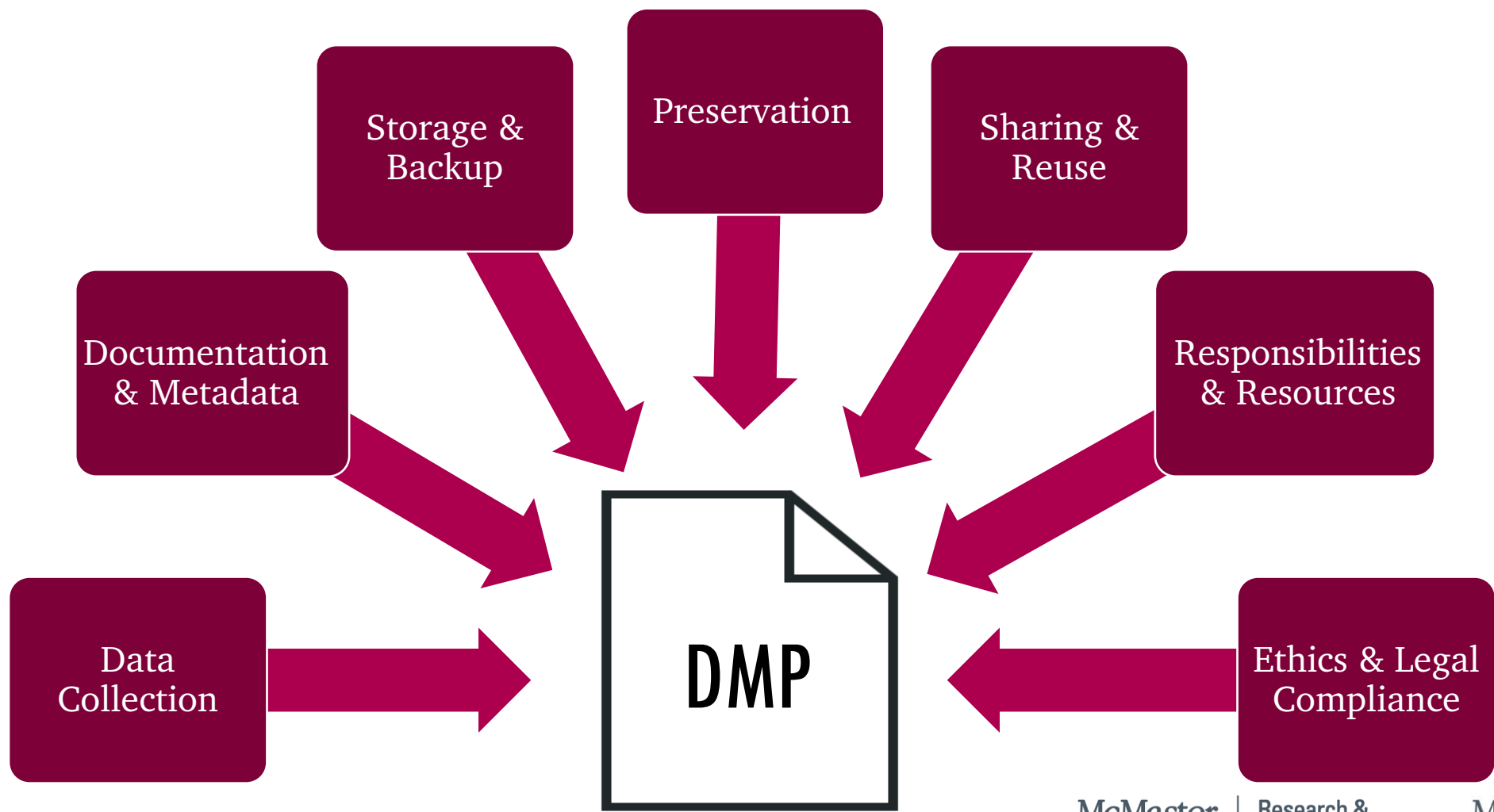
A **Data Management Plan (DMP)** is a living document describing your plan for how you will create, store, organize, document, secure, preserve, and share your research data.

- A document which speaks to the management of data both **during** the active phases of research and **after** the completion of the research project.
- DMPs guide researchers in articulating their plans for managing data; they do not set standards for what constitutes acceptable RDM practices

WHY SHOULD RESEARCHERS CREATE DMPS?

- Set out consistent strategies prior to starting your research for how data will be managed
- Identify the strengths & weaknesses in your current practices and decide how to integrate effective data management practices into your research
- An excellent way to engage partners and collaborators in ongoing conversation about how to best manage research data
- Many research funders require grant applicants to submit a DMP – Most UK & EU funders, NSF, NIH, Wellcome Trust, Tri-Agency (starting 2022)

WHAT GOES IN A DATA MANAGEMENT PLAN



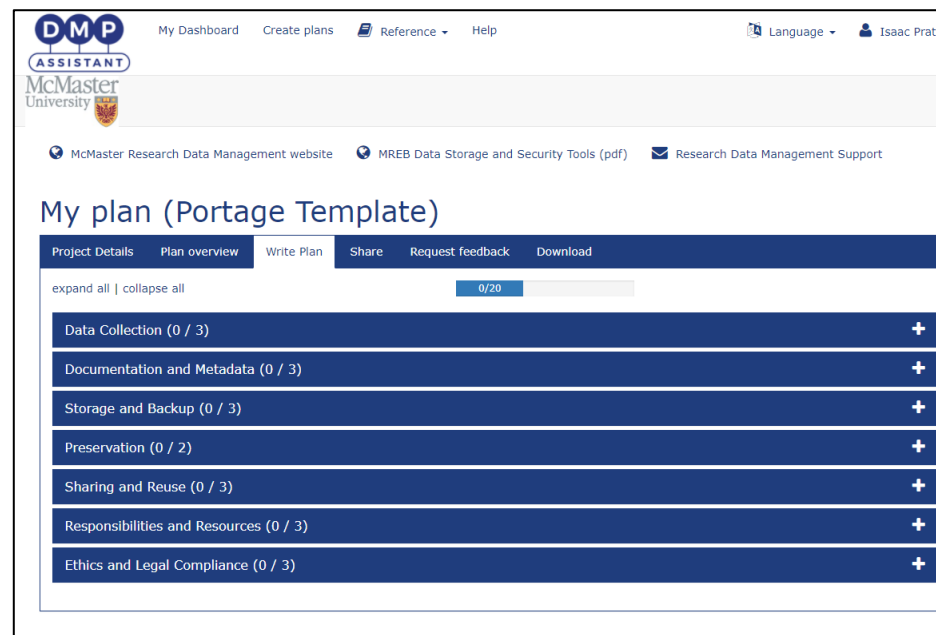
WHAT MAKES A GOOD DMP?

Portage Network has published a number of DMP exemplars available here: <https://portagenetwork.ca/tools-and-resources/training-resources/>

Some Rubrics exist from other jurisdictions:

- USA <https://osf.io/qh6ad/>
- UK <https://research-data-network.readme.io/docs/compliance-tools>
- UK Slides on ['How to Review DMPs'](#)

- A web-based, bilingual data management planning tool
- Available to all researchers in Canada
- A guide for best practices in data stewardship
- Exportable data management plans



<https://assistant.portagenetwork.ca/>

WHAT MAKES A GOOD DMP?

There are no absolute right answers to any of the questions but a good DMP:

- Has answers that are specific and detailed
- Shows that the researcher has given the topics due consideration and engaged with the issues
- Follows disciplinary best practices and norms
- Uses institutional resources and has sought advice where needed
- Properly justifies costs and restrictions
- Seems feasible to follow

WHAT MAKES A GOOD DMP?

Is the information specific enough?

- *“we will use suitable formats to ensure that our data can be preserved and sustained over the long term”*
- Which formats? Name them!
- Preserved where? Sustained how?

WHAT MAKES A GOOD DMP?

A better response...

- *“We will provide MP3 audio files for online dissemination. While this is not an open format, it is well-established and the most widely supported. High-resolution WAV files will be used for the archival master recordings.”*
- Is clear, specific and detailed
- Justify decisions

“BUILDING A DATA MANAGEMENT PLAN” WEBINAR

Check out the recording of my webinar online at:

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

A promotional graphic for a webinar. It features a laptop in the foreground with a hand typing on the keyboard. The background is white with a light blue circle and a green circle. The text is arranged as follows: 'Research Data Management Summer Series' at the top left, 'Building a Data Management Plan for your research project' in large maroon font in the center left, 'May 6, 2021 | 2:00 pm' below it, and 'Synchronous Workshop' in a green circle in the center right. Logos for McMaster University, Research & High Performance Computing, and the Sherman Centre are at the bottom.

Research Data Management Summer Series

Building a Data Management Plan for your research project

May 6, 2021 | 2:00 pm

Synchronous Workshop

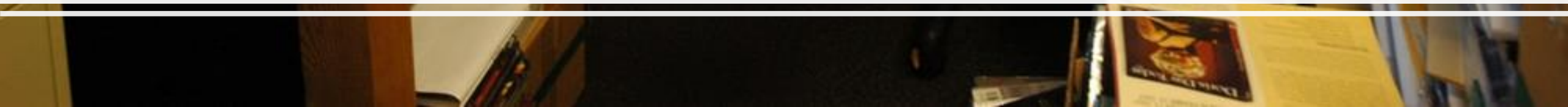
McMaster University | Research & High Performance Computing

McMaster University | Library

Lewis & Ruth Sherman Centre for Digital Scholarship



DATA MANAGEMENT PRACTICES



DATA ORGANIZATION

“A place for everything and everything in its place.”

- We recommend to researchers that they implement structured folder organization and file naming schemes.
- Good file names makes it easy to find and understand data and other files.
- There is no one right way to organize files, but the key to a good system is that it should be **descriptive, standardized, and consistently** implemented.

DATA STORAGE & BACKUP

“There are two types of data: data that is backed up and data that just hasn’t been lost yet”

A good data storage plan needs to balance **accessibility** and **convenience** against **security** and **reliability**. Different datasets have different needs and the balance will be different for each one.

3-2-1 Backup Strategy:


- 3 copies of your data where
- 2 copies each in a different storage system
- 1 copy is in a trusted off site location

RESEARCH DATA STORAGE FINDER TOOL

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

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

SELECT ALL **CLEAR SELECTIONS**

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

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

“STRATEGIES FOR RESEARCH DATA STORAGE AND BACKUP” WEBINAR

Check out the recording of my webinar online at:
<https://scds.github.io/intro-rdm/storage.html>

A promotional graphic for a webinar. It features a hand holding a key with a USB drive attached, set against a background of overlapping circles in light blue and lime green. The text is arranged as follows:

Research Data Management Summer Series

**Strategies for
research data
storage and backup**

Synchronous
Workshop

June 3, 2021 | 2:00 pm

Logos at the bottom include: McMaster University, Research & High Performance Computing, McMaster University Library, and the Lewis & Ruth Sherman Centre for Digital Scholarship.



DATA DEPOSIT

PUBLISHING DATA

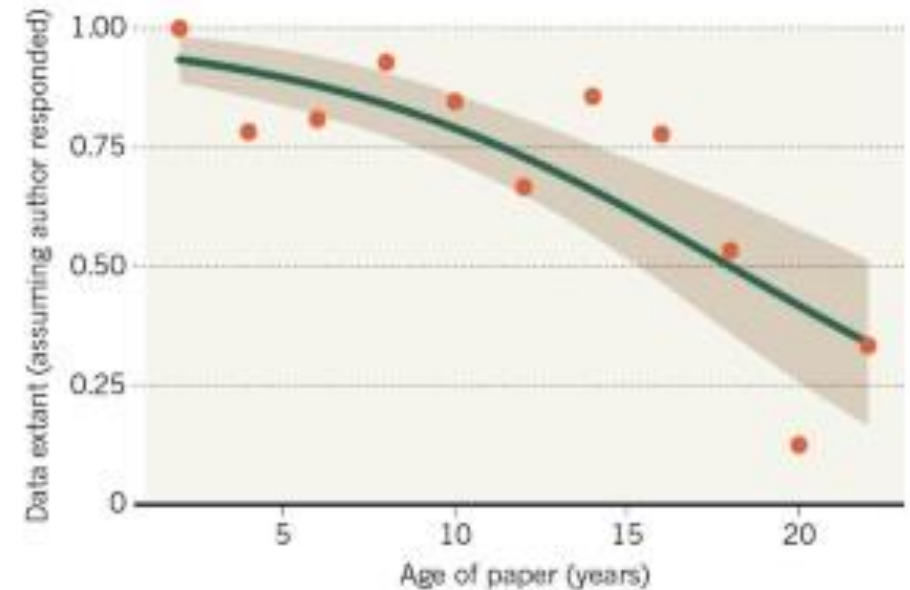
Tri-Agency policy will require that most research data be deposited in an online repository.

A 2013 paper showed that up to **80%** of data becomes unavailable after 20 years.

Vines et al 2013 <https://doi.org/10.1016/j.cub.2013.11.014>

MISSING DATA

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



THE VALUE OF DATA SHARING

Improve the **quality** of your research

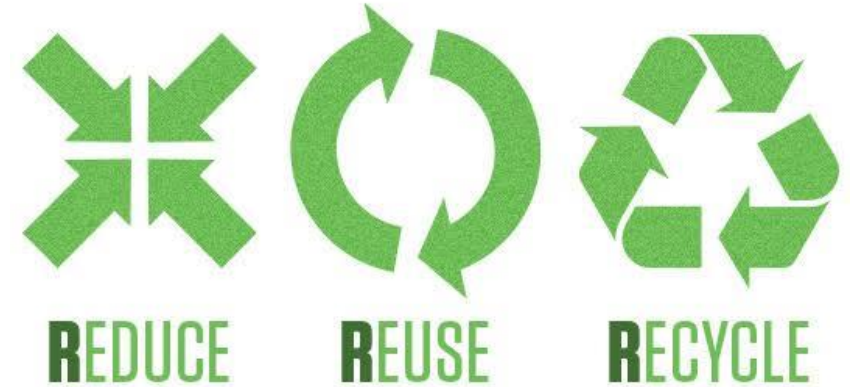
- Allow verification of results/code by peers
- Potential of ‘mega’ datasets

Improve the **impact** of your work

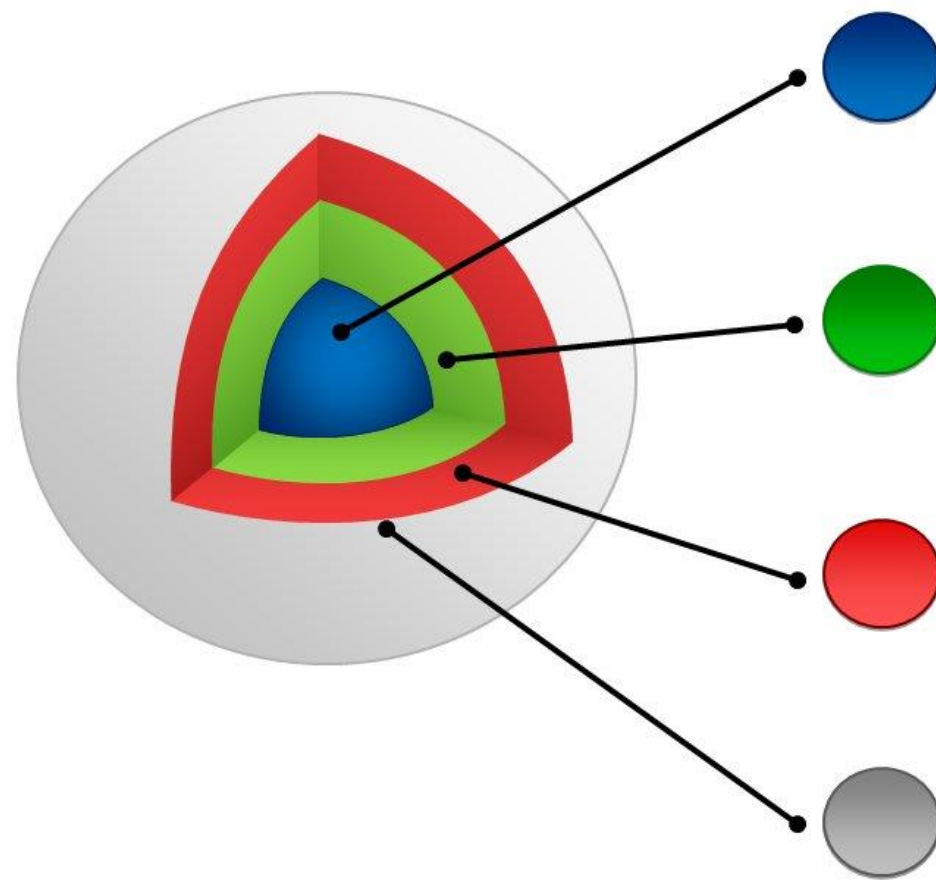
- Increases the potential visibility of research
- Can lead to new collaborations and partnerships
- Creates a lasting record of your work

Improve the **value** of your research

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



FAIR DATA



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

Findable
Accessible
Interoperable
Reusable

DATA DOCUMENTATION

Data documentation is like a Rosetta stone – it helps us understand the data.

We recommend researchers document their data (and their code) using **readme** files, **codebooks**, and **data dictionaries**

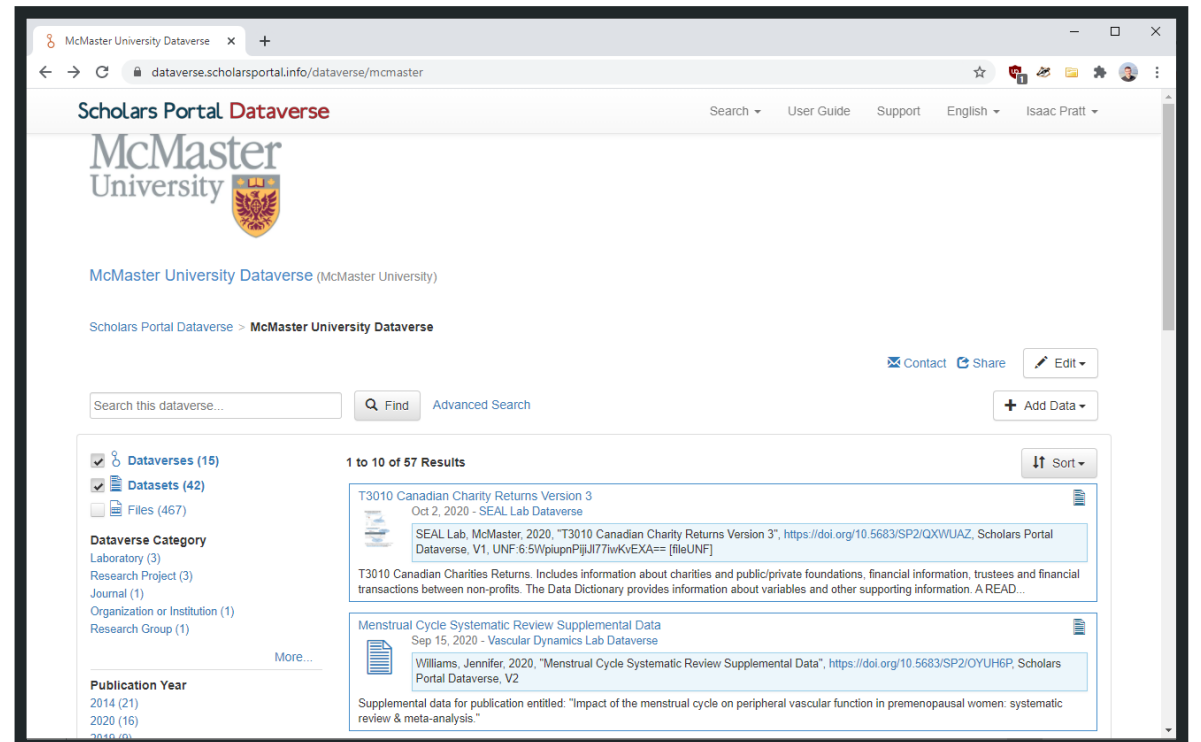
- A **readme** file is a simple document that describes the contents and organization of the dataset files.
- A **data dictionary** or **codebook** is a document describing the data and its variables

As with organization, the most important aspect of documentation is doing it.

MCMMASTER DATAVERSE

McMaster Dataverse is our Institutional Data Repository
<https://dataverse.scholarsportal.info/dataverse/mcmaster>

- Built for datasets
- Contains tools for tabular data exploration and analysis
- Allows researchers to control how they license and share their datasets



FEDERATED RESEARCH DATA REPOSITORY (FRDR)

- Built for large (1 TB+) datasets
- Datasets are actively reviewed by FRDR staff
- Datasets must be fully open but can be embargoed for a one year period



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

EXTERNAL DATA REPOSITORIES

Domain specific:

<https://www.nature.com/sdata/policies/repositories>

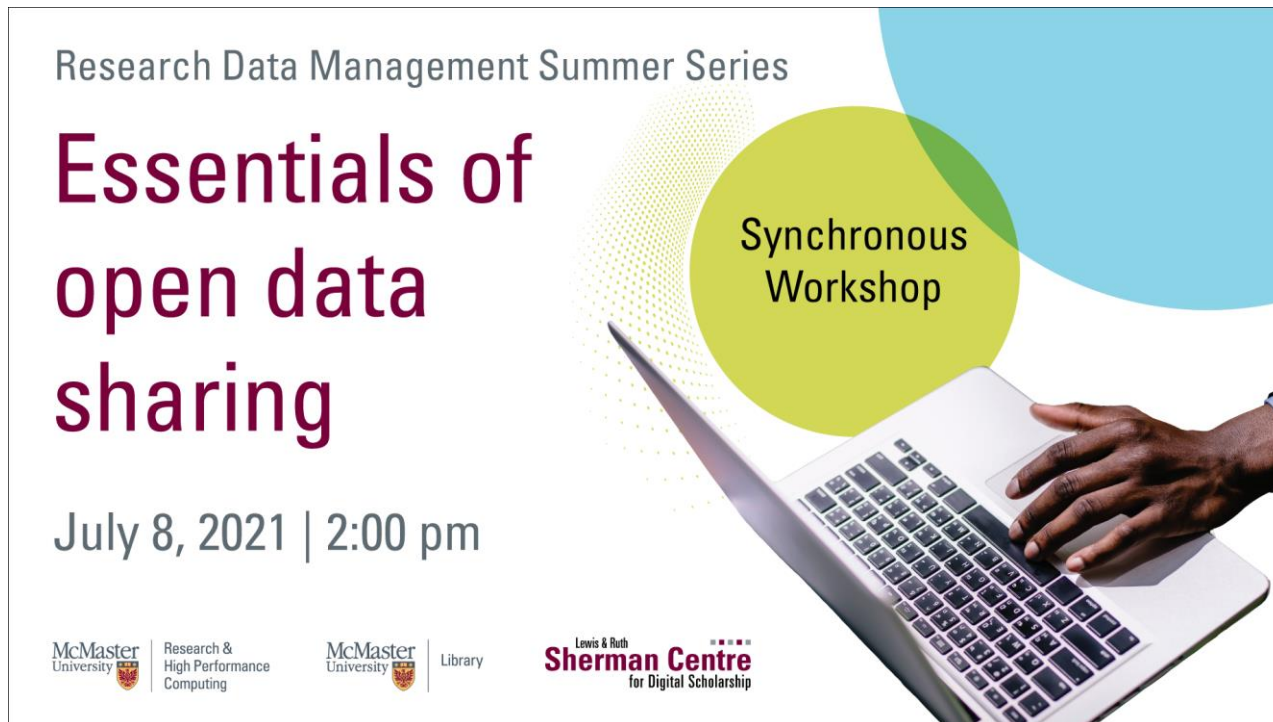
General: Zenodo, Figshare, Mendeley Data, etc

Search for repositories on re3data



“ESSENTIALS OF OPEN DATA SHARING” WEBINAR

Check out my next webinar on **Thursday, July 8, 2021 at 2 pm:**
<https://libcal.mcmaster.ca/event/3604377>



Research Data Management Summer Series

Essentials of open data sharing

Synchronous Workshop

July 8, 2021 | 2:00 pm

McMaster University | Research & High Performance Computing

McMaster University | Library

Lewis & Ruth Sherman Centre for Digital Scholarship

THANK YOU!

For more information:

Visit: library.mcmaster.ca/services/rdm

Check out the whole webinar series at:
<https://scds.github.io/intro-rdm/>

Contact us at: rdm@mcmaster.ca

RDM
@McMaster