



# Managing your research data

---

Introduction to the DMP Assistant

Isaac Pratt ([pratti@mcmaster.ca](mailto:pratti@mcmaster.ca)) & Jay Brodeur ([brodeujj@mcmaster.ca](mailto:brodeujj@mcmaster.ca))

McMaster University Library/RHPCS

Family Medicine Skill Builder Series | December 1, 2020

# Overview

- ✧ A brief Research Data Management primer (5 mins)
- ✧ Introduction to data management plans (DMPs) (10 mins)
- ✧ Portage DMP Assistant (35 mins)
  - ✧ Demo (5 mins)
  - ✧ Discussion & Questions (30 mins)
- ✧ Where to go for help (5 mins)

# Learning objectives

By the end of this session you should be able to:

- ✧ **Describe** the basic activities and considerations of research data management (RDM)
- ✧ **Explain** the purpose and value of a Data Management Plan (DMP)
- ✧ **Use** the Portage DMP Assistant to draft a DMP
- ✧ **Apply** your knowledge and experience to **create** a DMP and/or
- ✧ **Identify** additional resources and support to help you complete one

# Take a moment to think about the following...

- ✧ If you were asked to share your data with another researcher would they be able to make sense of your work?
- ✧ If you needed to locate your data files from 5 years ago, how easy would they be to find and use?

# Applying RDM best practices will benefit...

## **Researchers and their collaborators**

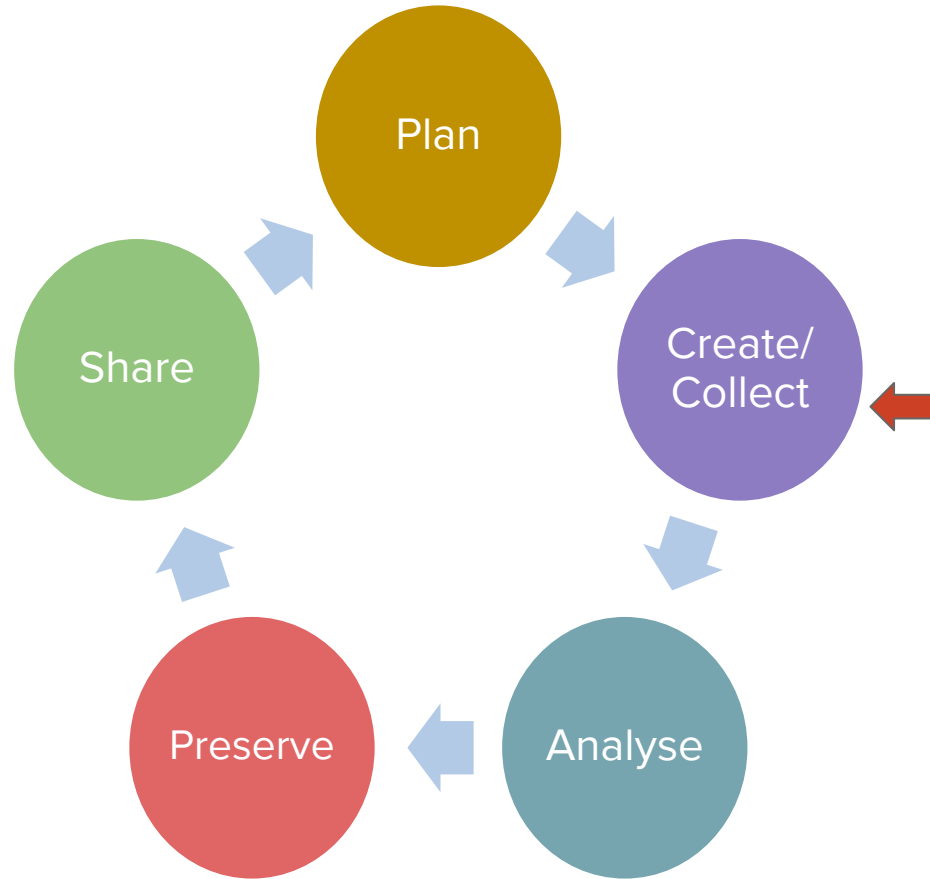
- ✧ Improves research efficiency and productivity
- ✧ Provides extra credit for research work
- ✧ Increases research impact
- ✧ May help to meet funding requirements

## **Research Communities**

- ✧ Accelerates discovery
- ✧ Enables validation and verification

## **Funders, governments and the public**

- ✧ Improves return on investment
- ✧ Increases research transparency
- ✧ Data as a public good



## Research Data Management

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

Data management plans

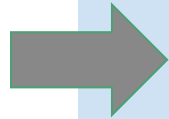
Provisioning storage, backup

Hardware, software, systems

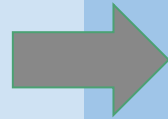
Ethics, IP, commercial reqs

Personnel, support, training

Data in  
Planning



Data in  
Motion



Data  
at Rest

Data management plans

Provisioning storage, backup

Hardware, software, systems

Ethics, IP, commercial reqs

Personnel, support, training

Collection & organization

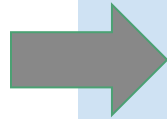
Storage, backup, recovery

Documentation & metadata

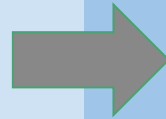
Access & Collaboration

Analysis

Data in  
Planning



Data in  
Motion



Data  
at Rest



Data management plans

Provisioning storage, backup

Hardware, software, systems

Ethics, IP, commercial reqs

Personnel, support, training

Collection & organization

Storage, backup, recovery

Documentation & metadata

Access & Collaboration

Analysis

Data/code archiving

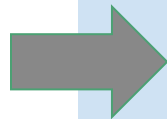
Curation, review, revision

Data/code preservation

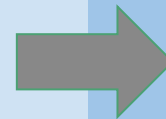
PIDs and linking

Data/code licensing, sharing

Data in  
Planning



Data in  
Motion



Data  
at Rest

Data management plans

Provisioning storage, backup

Hardware, software, systems

Ethics, IP, commercial reqs

Personnel, support, training

Collection & organization

Storage, backup, recovery

Documentation & metadata

Access & Collaboration

Analysis

Data/code archiving

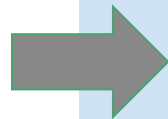
Curation, review, revision

Data/code preservation

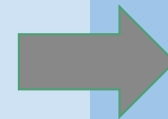
PIDs and linking

Data/code licensing, sharing

Data in  
Planning



Data in  
Motion



Data  
at Rest



# Where can I learn more and find help?

- ✧ Library Research Data Management website:

[library.mcmaster.ca/services/rdm](https://library.mcmaster.ca/services/rdm)

**RDM**  
@McMaster

- ✧ Email us for a research data management consultation

[rdmgmt@mcmaster.ca](mailto:rdmgmt@mcmaster.ca)

# Where can I learn more and find help?

- ✧ Portage Network <https://portagenetwork.ca/>
- ✧ A Canadian network of experts in Data Management working to build capacity and develop services and infrastructure in Canada
  - ✧ Portage's [Brief Guide to Research Data Management Guide](#)



## A Brief Guide: Research Data Management

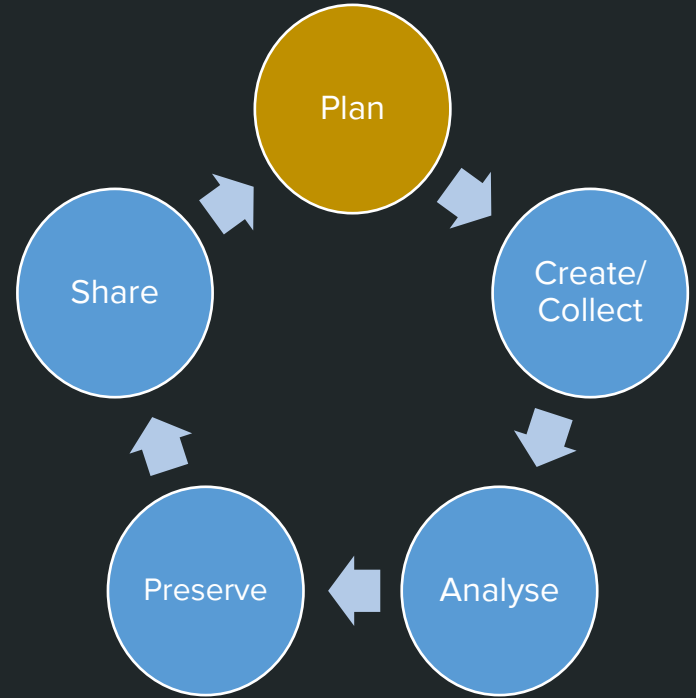
This brief guide presents a set of good data management practices that researchers can adopt, regardless of their data management skills and levels of expertise.

<p style="text-align: center;"><b>1</b></p> <p style="text-align: center;"><i>Save your raw data in original format</i></p> <ul style="list-style-type: none"> <li>1.1 Don't overwrite your original data with a cleaned version.</li> <li>1.2 Protect your original data by locking them or making them read-only.</li> <li>1.3 Refer to this original data if things go wrong (as they often do).</li> </ul>	<p style="text-align: center;"><b>2</b></p> <p style="text-align: center;"><i>Backup your data</i></p> <ul style="list-style-type: none"> <li>2.1 Use the <b>3-2-1 rule</b>: Save three copies of your data, on two different storage mediums, and one copy off site.</li> <li>2.2 Do not backup or store sensitive data on a commercial cloud (Dropbox/Google Drive, etc.).</li> </ul>
--	---

# The RDM Landscape

---

- Tri-Agency requirements
- Resources





### Partnerships and Projects

- Meet Canada's scientists and professionals
- Canadian High Arctic Research Station
- Northern Contaminants Program
- Canada's Extended Continental Shelf Program
- Health Canada: Centre for Vaccine Evaluation
- Science and Technology Data
- Research Funding Collaboration
  - Co-operative Funding Programs
  - Policies and Guidelines
    - Open Access
    - Selecting the Appropriate Federal Granting Agency
    - Research Ethics
    - Responsible Conduct of Research
    - Public Communications

## Tri-Agency Statement of Principles on Digital Data Management

### 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.

As publicly funded organizations, the agencies are strong advocates for making the results of the research they fund as accessible as possible. In promoting access to research results, they aspire to advance knowledge, avoid research duplication and encourage reuse, maximize research benefits to Canadians and showcase the accomplishments of Canadian researchers. These aspirations align with the Government of Canada's commitment to open science, as described in [Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation](#)(2014).

Research data include observations about the world that are used as primary sources to support scientific and technical inquiry, scholarship and research-creation, and as evidence in the research process.<sup>1</sup> Research data are gathered through a variety of methods, including experimentation, analysis, sampling and repurposing of existing data. They are increasingly produced or translated into digital formats. When properly managed and responsibly shared, these digital resources enable researchers to ask new questions, pursue novel research programs, test alternative hypotheses, deploy innovative methodologies and collaborate across geographic and disciplinary boundaries. The ability to store, access, reuse and build upon digital research data has become critical to the advancement of science and scholarship, supports innovative solutions to economic and social challenges, and holds tremendous potential for Canada's productivity, competitiveness and quality of life.

Tri-Agency Statement of Principles on Digital Data Management

<http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>

# Draft Tri-Agency RDM Policy

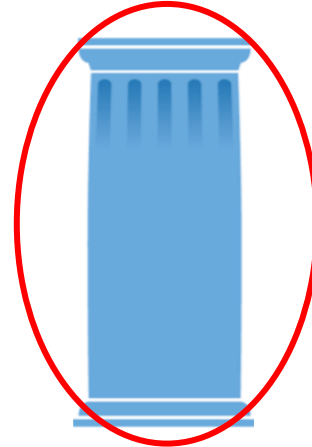
Tri-Agency [draft data management policy](#)<sup>1</sup> — June, 2018

✧ Will apply to grant recipients and institutions administering tri-agency funds.

**Institutional  
Strategy**



**DMPs**



**Deposit**



Phased and incremental implementation

<sup>1</sup>[http://www.science.gc.ca/eic/site/063.nsf/eng/h\\_97610.html](http://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html) [shortened: <https://bit.ly/2OsJB8w>]

# Existing Tri-Agency policy

## CIHR:

- ✧ Some publication-related research data including bioinformatics, atomic, and molecular coordinate data must be deposited into a public database immediately upon publication of research results.
- ✧ Data sets must be retained for a minimum of 5 years after the end of a grant.

## SSHRC:

- ✧ All research data collected with the use of SSHRC funds must be preserved and made available for use by others within a reasonable period of time (2 years after project completion).
- ✧ researchers have an obligation to preserve and safeguard the data.



# Scholars Portal Dataverse



<https://www.youtube.com/watch?v=UDFGqRY61fQ>

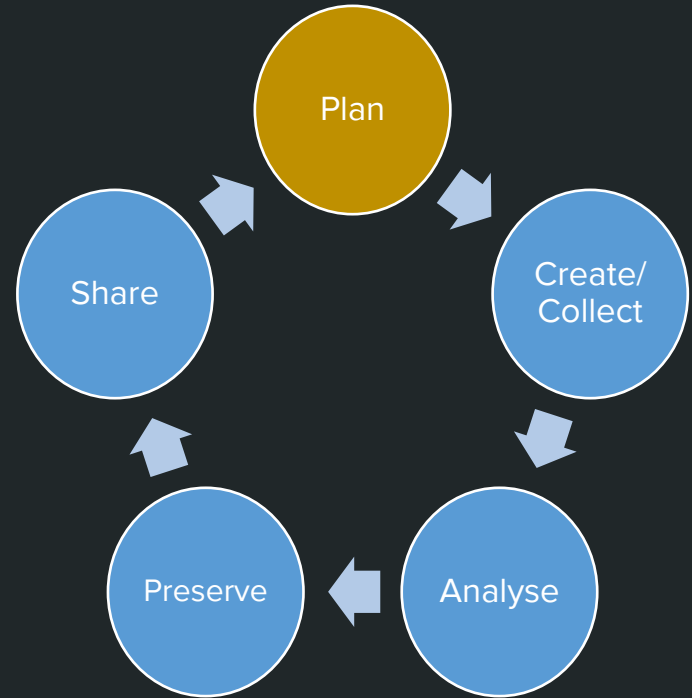
- ✧ A data repository for researchers at Ontario's universities -- **free** and **open** for all researchers in Canada
- ✧ An online platform to share, preserve, cite, explore and analyze research data.
- ✧ Allows researchers to control how they share their data.
- ✧ Locally administered at McMaster with servers in Canada.

<http://dataverse.scholarsportal.info>

# RDM Planning

---

- Considerations
- DMP Assistant



# Data Management Plans

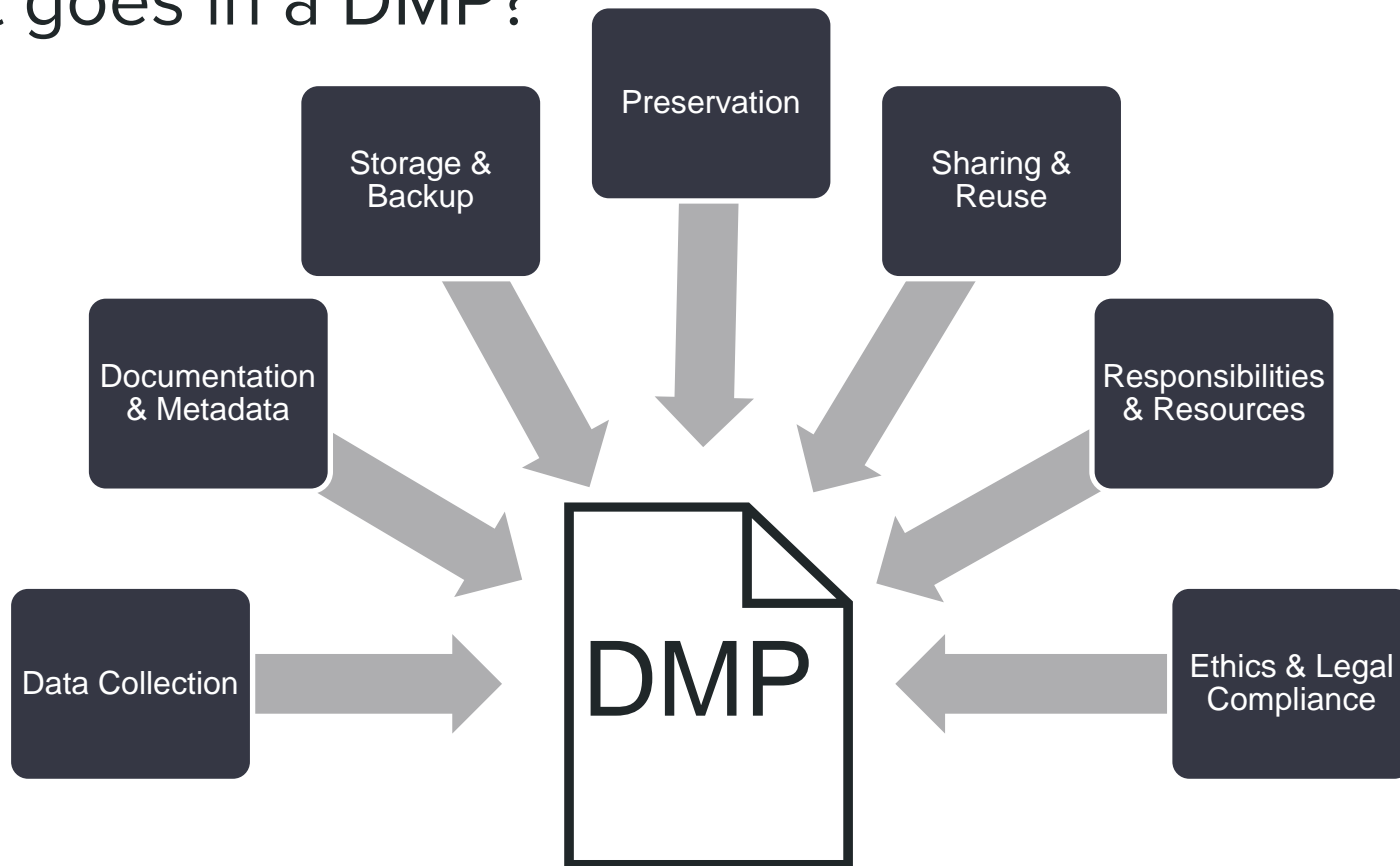
DMPs *are living documents* that should outline:

- how data will be collected, documented, formatted, protected & preserved;
- how existing datasets will be used and what new data will be created
- whether and how data will be shared; and where data will be deposited.
- roles & responsibilities for managing, including succession plans
- ethical, legal and commercial constraints of the data
- methodological considerations that support or preclude data sharing

Grant applicants are encouraged to use standardized tools to develop their DMPs, such as the **Portage Network's DMP Assistant**<sup>1</sup>

[1] <https://assistant.portagenetwork.ca>

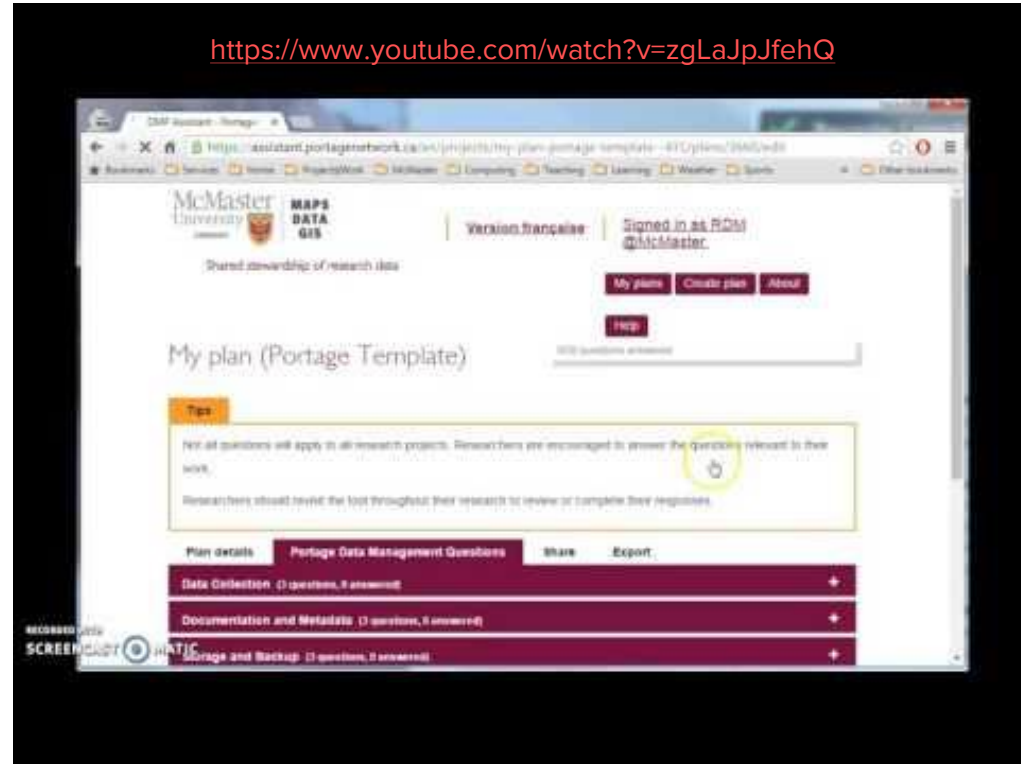
# What goes in a DMP?



# Portage DMP Assistant

- ❖ 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://www.youtube.com/watch?v=zgLaJpJfehQ>



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

# Activity - Develop a DMP (~35 minutes)

1. Navigate to [assistant.portagenetwork.ca](https://assistant.portagenetwork.ca)
  - Create an account & login (check email for verification)
  - Or login with user: **mcm.rdm.demo@gmail.com** | pass: **demopass**
  - Or download list of questions at: [u.mcmaster.ca/portage-dmp-template](https://u.mcmaster.ca/portage-dmp-template)
2. Open our session's padlet: [u.mcmaster.ca/dmp-padlet](https://u.mcmaster.ca/dmp-padlet)
3. Create a plan for a current, past, upcoming, or prospective research project
  - Note points of interest, difficult or unanswered questions, or other thoughts on the padlet

# Reflections and Discussion

## What are your reflections on:

- ✧ Which sections / questions were particularly challenging? How so?
- ✧ What kind of support or resources would help you complete this plan?
- ✧ The overall experience?
- ✧ The interface and its content?

# Thank You.

For more information:



[library.mcmaster.ca/services/rdm](https://library.mcmaster.ca/services/rdm)  
[rdmgmt@mcmaster.ca](mailto:rdmgmt@mcmaster.ca)

McMaster Library's RDM webpage  
and primary contact



[portagenetwork.ca/](https://portagenetwork.ca/)

Portage Network page: Access to RDM  
information and DMP Assistant