

Building a **Data Management Plan** for your research project.

Isaac Pratt, PhD + Danica Evering, MA

November 16, 2022





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.

WabbitWanderer, “Cockpit Island in Cootes Paradise, Hamilton Ontario, on an autumn afternoon,” October 28, 2018, Flickr - <https://www.flickr.com/photos/44251652@N08/44101709870/in/photostream/>

Code of Conduct

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

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

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

Session Recording and Privacy

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

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

Certificate Program

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

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

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

Hello! A bit about us:

We are Research Data Management Specialists

Isaac Pratt, PhD

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

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

Danica Evering, MA

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

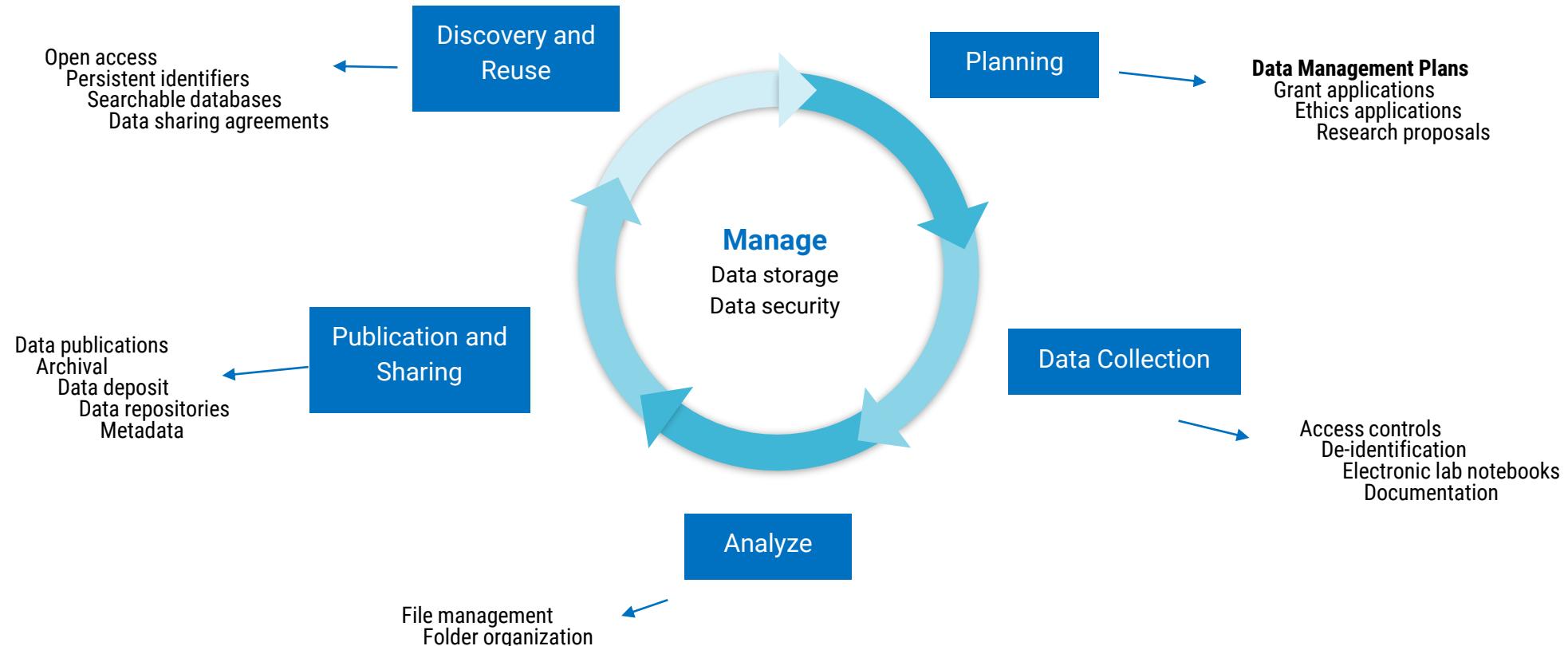
I have an MA in **Media Studies** from Concordia University.

Outline

-  What is RDM and why is it important?
-  Data Management Plans (DMPs)
 - Why are they important?
 - What goes in one?
 - What makes a good plan?
-  Funder Requirements for DMPs
-  Digital Research Alliance of Canada DMP Assistant tool

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**.



For a fuller introduction to RDM see our earlier webinar “Best Practices for Managing Data in your Research” - <https://scds.github.io/intro-rdm/intro.html>

Why is RDM important? **Data** are valuable.



Why is Research Data Management (RDM) important?

Research Data Management best practices	make your research better.
Proper data organization and planning ahead	saves time and resources.
Good data storage and backup strategies	help avoid loss of data from theft, corruption, or damage to storage devices.
Sharing data openly	allows others to reproduce and verify research results.
Depositing and Publishing data	increases the visibility of research and citations.

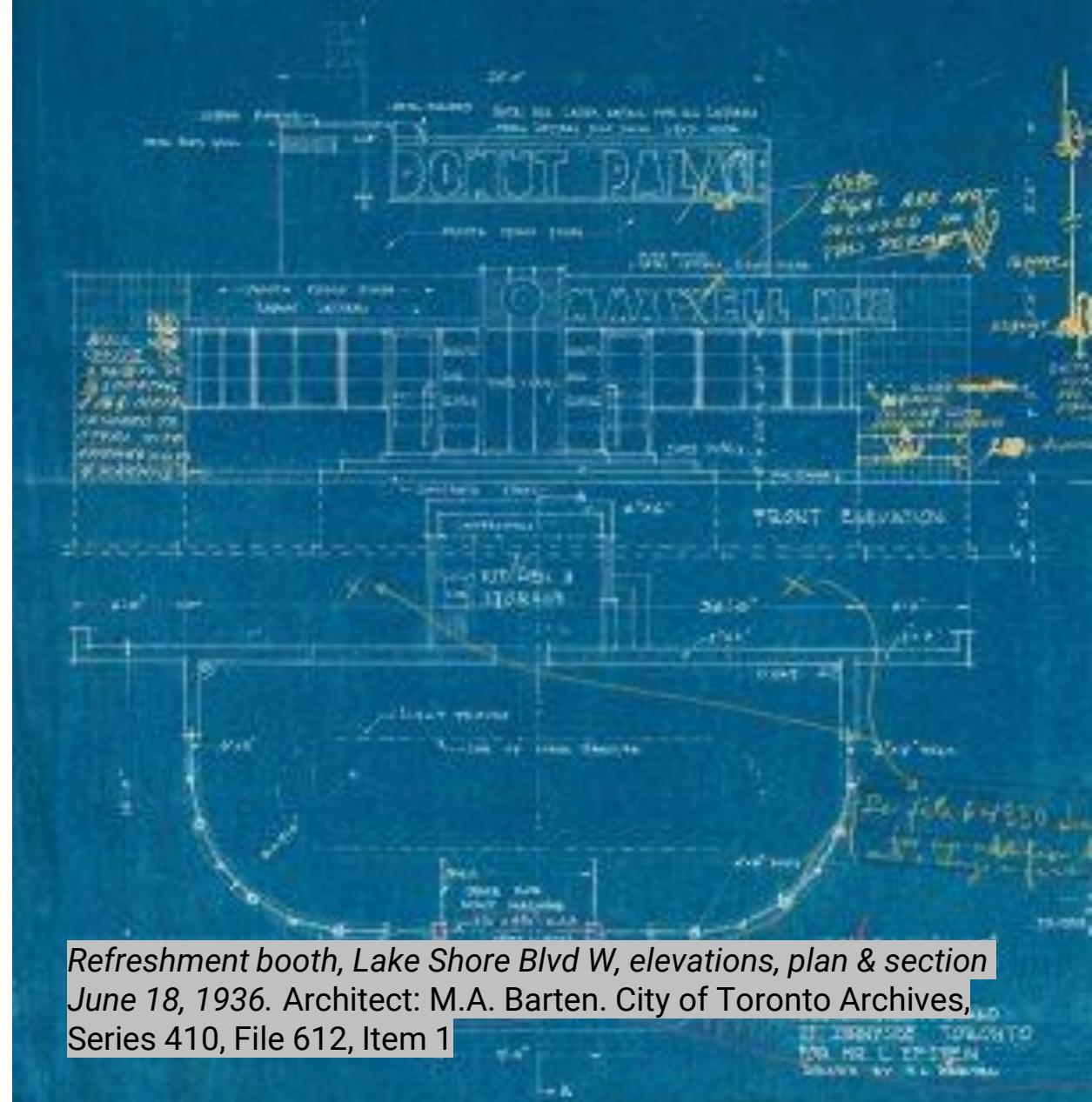
Photo by Louis Reed on Unsplash.

Data Management Plans:

Set up a system for best practices for your project.

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.

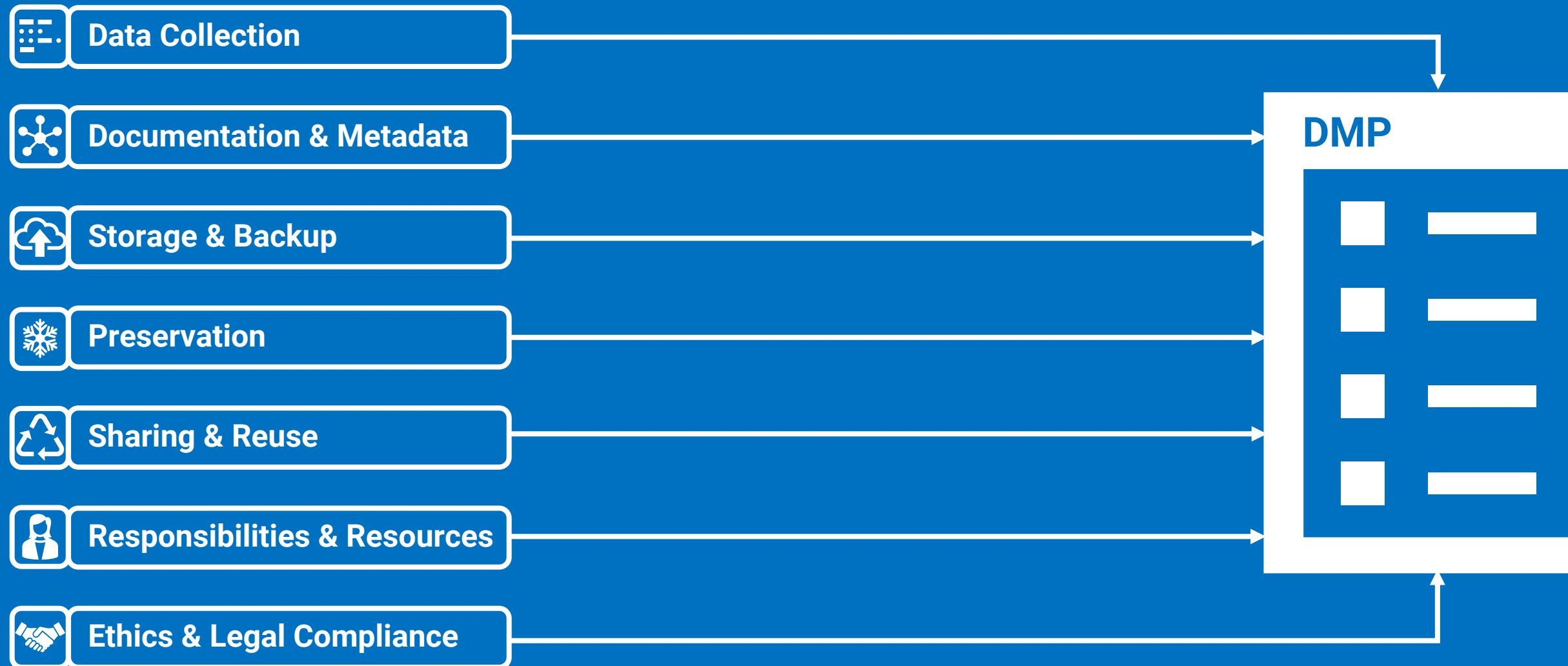
- Creating your own DMP is straightforward using web tools such as [**DMP Assistant**](#).
- DMPs ask pointed questions of researchers to help them **articulate their plans for managing data**; *they do not compel researchers to manage data differently*.
- DMPs outline how you will manage data both **during** the active phases of your research and **after** the completion of the research project.



Refreshment booth, Lake Shore Blvd W, elevations, plan & section
June 18, 1936. Architect: M.A. Barten. City of Toronto Archives,
Series 410, File 612, Item 1



What goes in a Data Management Plan?



Bad management makes data **vulnerable**, **messy**, + **disconnected**.

Consider the most common approach to data management:

- **Storage:** Data is stored on laptop or desktop hard drives and backed up to a collection of miscellaneous external hard drives accumulated over the years.
- **Documentation:** Data is not consistently documented
- **Sharing:** Data is not published or shared outside the research group except by direct request.

This approach is **vulnerable** to data loss and makes working with the data frustrating.

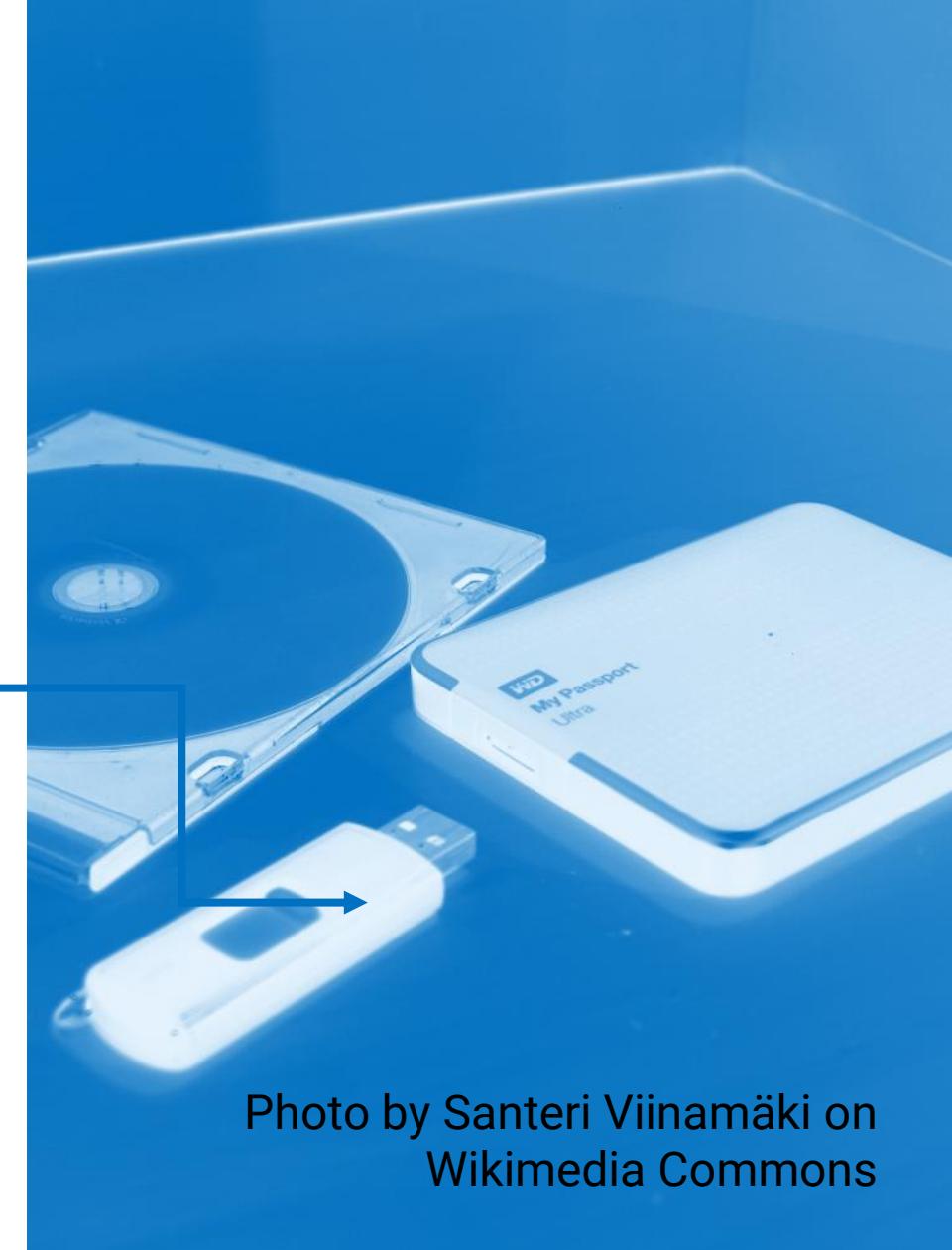


Photo by Santeri Viinamäki on
Wikimedia Commons

Why should I create a Data Management Plan?



Set out consistent strategies **prior to starting your research** for how data will be managed, shared, and archived.



Identify the **strengths & weaknesses** in your current practices.



Ensure **quality assurance** and decide how to integrate effective data management practices into your research.



Make sure your data stays safe and align with **ethical responsibilities**.



Photo by National Cancer Institute
on Unsplash.

DMPs support collaboration within your lab or research team.

- Engage research partners and collaborators in ongoing conversation about how to best manage research data.
- Establish and consistently lay out data practices for a lab.
- Set up storage and security systems, with timelines for backups and updates.
- Ensure contingency plans and responsibilities for unexpected events – illness, moving universities, ransomware attack.



Data Management Plans are “living documents”

- A living document reflects the inevitability of change.
- A living document is edited and updated on an ongoing basis.
- Update your DMP as your project evolves and consider reviewing it along with your password check every 3 months.
- Approached this way, a DMP can be a very useful research tool!



Waves breaking at the Fox Island boat launch with Shaws Cove in the background, animated.
By Thorsten Lindner on Wikimedia Commons



Tri-Agency RDM Policy 2021

Government of Canada Gouvernement du Canada

Search Canada.ca

MENU ▾

Home > Interagency research funding > Policies and Guidelines > Research Data Management

Research Data Management

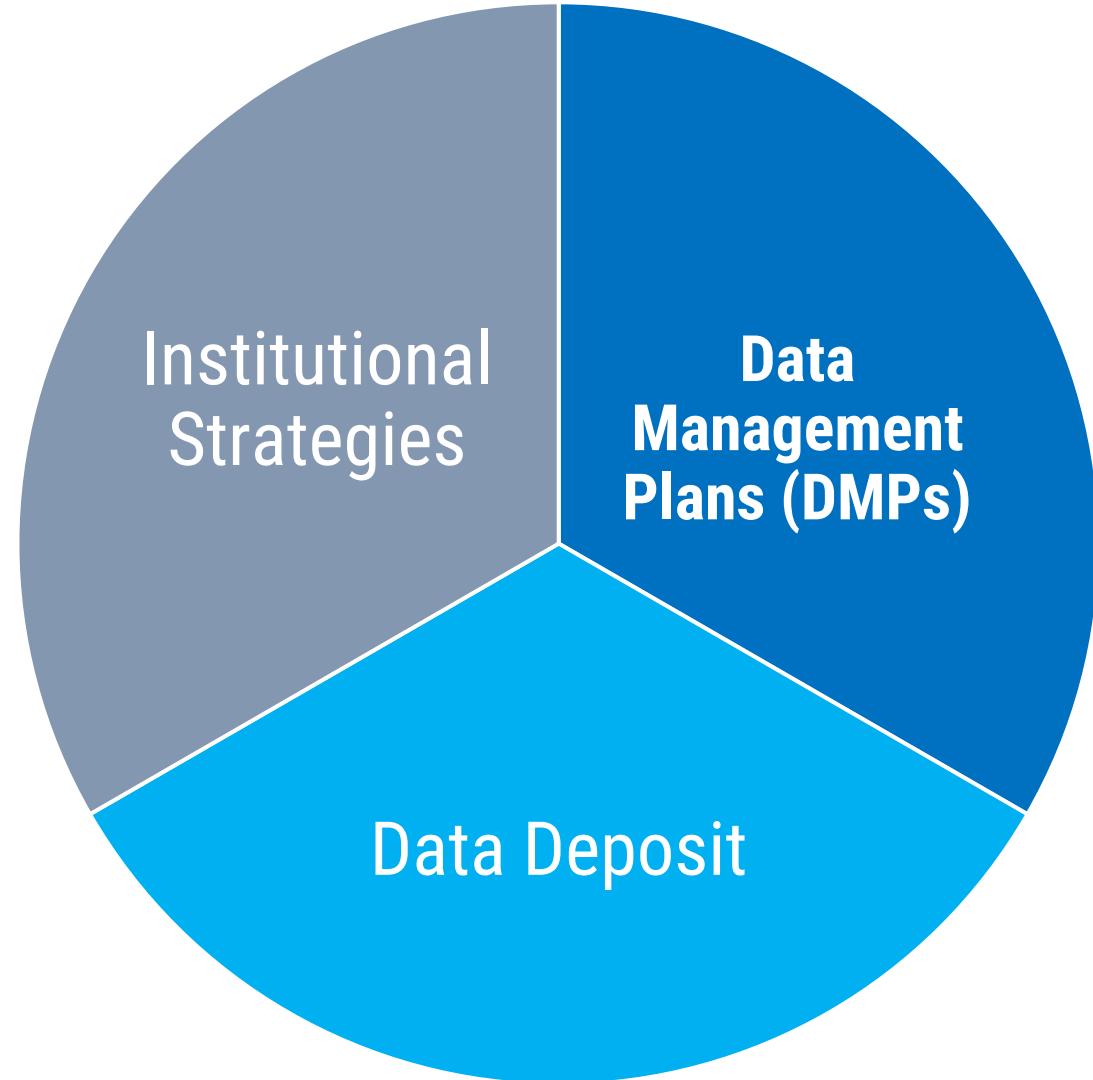
Tri-Agency Statement of Principles on Digital Data Management

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



“All grant proposals submitted to the agencies should include methodologies that reflect best practices in RDM. **For certain funding opportunities, the agencies will require data management plans (DMPs) to be submitted to the appropriate agency at the time of application**, as outlined in the call for proposals; in these cases, the DMPs will be considered in the adjudication process.

DMPs are living documents that can be modified to accommodate changes throughout the course of a research project.”

Innovation, Science and Economic Development Canada. “Tri-Agency Research Data Management Policy.” Government of Canada. Innovation, Science and Economic Development Canada, March 15, 2021.
<https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-research-data-management-policy>

Tri-Agency Data Management Plan Requirements:

Canadian Institutes of Health Research (CIHR)

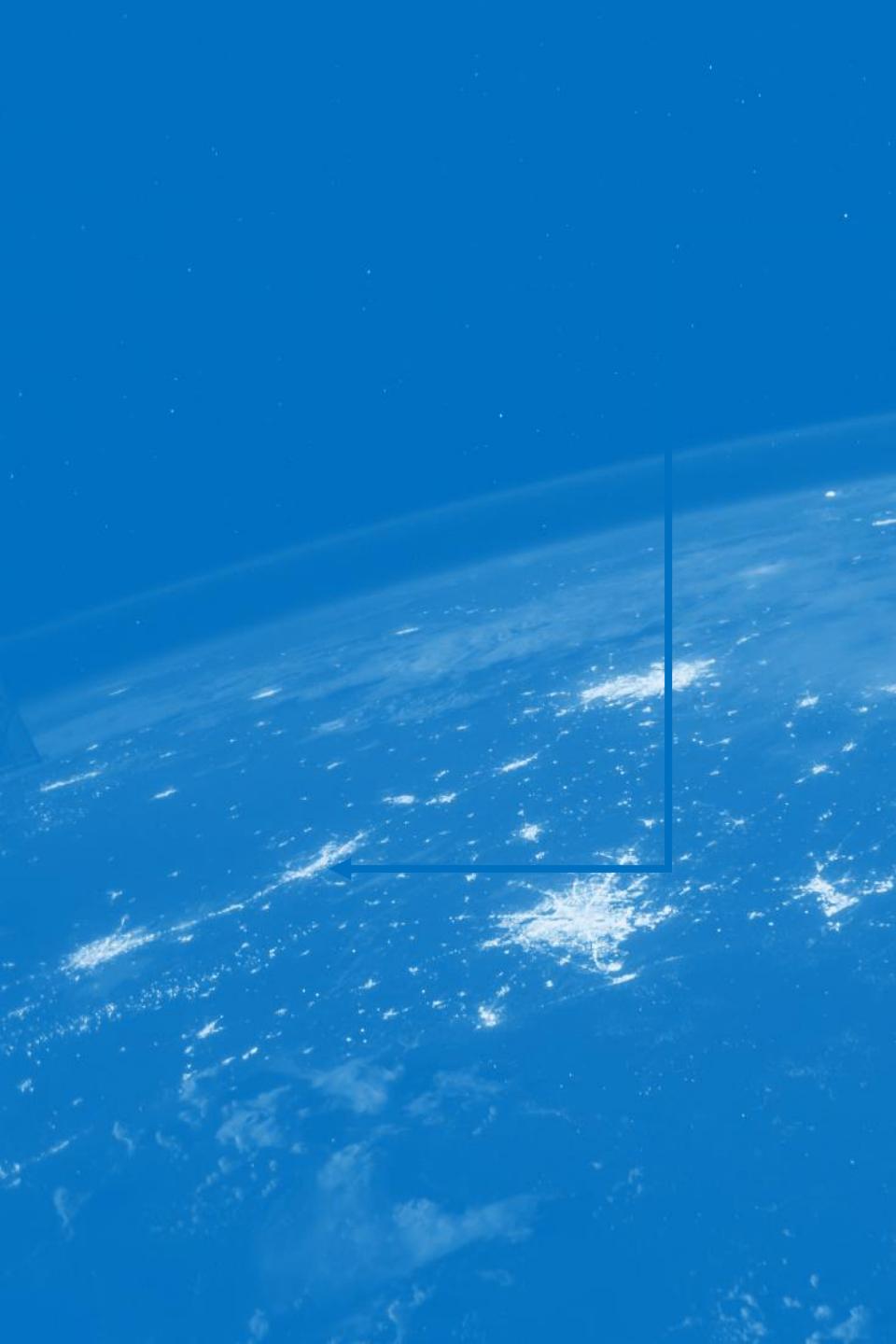
- Network Grants in Skin Health and Muscular Dystrophy (Anticipated launch fall 2022 or early winter 2023)
- Virtual Care/Digital Health Team Grants (Anticipated launch fall 2022 or early winter 2023)
- Data Science for Equity (Anticipated launch fall 2022 or early winter 2023)

Natural Sciences and Engineering Research Council of Canada (NSERC)

- Subatomic Physics Discovery Grants - Individual and Project (Anticipated launch summer 2023)

Social Sciences and Humanities Research Council of Canada (SSHRC)

- Partnership Grants Stage 2 (Anticipated launch summer 2023)



Data Management Plans align with global research funding.

- Most UK & EU funders, Wellcome Trust
- Other Canadian funders, such as the International Development Research Centre
- United States - National Science Foundation (NSF) Grants
- United States - National Institutes of Health (NIH) just released outline for their Data Management and Sharing Policy for rollout in 2023 – requiring a **Data Management and Sharing Plan (DMSP)**



DMP Exemplars

The Digital Research Alliance of Canada has several DMP exemplars, [available here](#).

We're going to look at the DMP for the “People, Places, Policies and Prospects: Affordable Rental Housing for Those in Greatest Need,” project.
<https://zenodo.org/record/4062466>

Catherine Leviten-Reid, Jasmine Hoover, Cape Breton University.



Data Management Plan for People, Places, Policies and Prospects (v. 1.0)

Data Collection

What types of data will you collect, create, link to, acquire and/or record?

Data collected during our projects may include, but are not limited to, those gathered from surveys, in-depth interviews, focus groups, community conversations and arts-based methods such as photography. This means we will potentially generate numeric, audio, image, video and text-based data.

What file formats will your data be collected in? Will these formats allow for data re-use, sharing and long-term access to the data?

This research project is collecting a variety of types of data. Examples of these include XML¹ and CSV² for databases and spreadsheets, JPG³ or TIFF⁴ files for images, MP3⁵ files for sound and TXT⁶ for text. Each of these file types are non-proprietary, ensuring ease and flexibility of reuse.

Providing definitions for acronyms in each new section or on each page prevents readers unfamiliar with the terminology from having to revisit previous sections.

Data Collection

- What types of data will you collect, create, link to, acquire and/or record?
- Where is your data coming from? Are you re-using existing data?
- What file formats will your data be collected in?
- How will your data be organized? SQL database, spreadsheet, etc
- What conventions and procedures will you use to structure, name and version-control your files?

Data Collection

Data collected during our projects may include, but are not limited to, those gathered from surveys, in-depth interviews, focus groups, community conversations and arts-based methods such as photography. This means we will potentially generate numeric, audio, image, video and text-based data.

This research project is collecting a variety of types of data. Examples of these include XML and CSV for databases and spreadsheets, JPG or TIFF files for images, MP3 files for sound and TXT for text. Each of these file types are non-proprietary, ensuring ease and flexibility of reuse.



Data Collection

All files will use a conventional **naming standard**. File names should include the **grant name** (in shortened form), a **summary** of the file's content, the **region** and the **date** (in the format YYYY/MM/DD).

An example is the following:

prospects_interviewguide_ON_20200617.

Document **versions** should be named sequentially (with file names ending in v1, v2 etc.).

An example is the following:

prospects_interviewguide_ON_20200617_v1.



Data Collection: Resources

- File formats:
 - DataverseNO [Prepare your data: Preferred file formats](https://site.uit.no/dataverseno/deposit/prepare/#what-are-preferred-file-formats)
<https://site.uit.no/dataverseno/deposit/prepare/#what-are-preferred-file-formats>
 - DCN Data Curation Primers on preserving different file formats
<https://datacurationnetwork.org/outputs/data-curation-primers/>
- McMaster RDM page on file naming and organization
 - <https://rdm.mcmaster.ca/organize#tab-file-folder-organization>

Documentation & Metadata

“The utility and longevity of data relates directly to how complete and comprehensive the metadata are.” Michener, 2015

- What documentation will be needed for the data to be read and interpreted correctly in the future?
- How will you make sure that documentation is created or captured consistently throughout your project?
- Will you use a metadata standard and/or tools to document and describe your data?

Documentation & Metadata

In order for data to be potentially reused, all data files should include a description of team members responsible for creating the data, how the data were collected, the **code book** (if involving survey data), the **interview guide** (if involving qualitative data), any issues affecting data quality and other pertinent background information which allows the content to be easily understood by others.

All files containing spreadsheets must include column names which are easily interpreted, even though they will be defined in a code book.



Documentation & Metadata

Team researchers engaged in data analysis using software will create **logs and syntax files** to ensure that the steps leading to the final results are documented and saved. No identifying information of participants may be included in data files. Metadata must also include the grant name and funders (SSHRC and CMHC).

Since these data files will be deposited in the Scholars Portal Cape Breton University Dataverse, the **Data Documentation Initiative (DDI) metadata standard** will be applied.



Documentation & Metadata: Resources

Data Documentation

- McMaster RDM page <https://rdm.mcmaster.ca/organize#tab-file-folder-organization>
- Cornell ReadMe template
<https://cornell.app.box.com/v/ReadmeTemplate>

Metadata

- Dataverse North Metadata Best Practices Guide v 2.0
<http://hdl.handle.net/2429/73609>
- DCC list of disciplinary metadata standards
<https://www.dcc.ac.uk/guidance/standards/metadata>



CODE: STARLING

RDM Certificate Program

- Certificate you can add to your CV or ORCID
- Attend 7 RDM workshops to receive a certificate!
- Go to this website to verify today's session:
<https://u.mcmaster.ca/verification>
- Learn more about the Certificate Program:
<https://scds.ca/certificate-program>



Storage & Backup

Data loss is more common than you think. Hard drives, USB drives, and other storage media can fail easily or be lost/stolen.

- What are the anticipated storage requirements for your project?
- How and where will your data be stored and backed up during your research project?
- How will the research team and other collaborators access, modify, and contribute data throughout the project?

Storage & Backup

Storage space is anticipated to be approximately **100 GB**. The data will be stored for **5 years locally**, with a permanent copy held in the Scholars Portal Cape Breton University **Dataverse**.

The **3-2-1 backup rule** will be followed for data storage and backup. All team members will upload their files to a cloud-based server located in Canada, to be identified by the project lead. Sensitive files are to be encrypted.

OneDrive is used to store, share, and work with data.



Storage & Backup: Resources

Data Storage:

- McMaster Research Data Storage Finder
<https://u.mcmaster.ca/storagefinder>

Backup:

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.)

1

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



Preservation

- Where will you deposit your data for long-term preservation and access at the end of your research project?
- Indicate how you will ensure your data is preservation ready.
 - Consider preservation-friendly file formats, ensuring file integrity, anonymization and de-identification, inclusion of supporting documentation.

Preservation

Data collected during this grant should normally be indexed/archived on the Scholars Portal Cape Breton University **Dataverse** in accordance with the SSHRC policy on data sharing. To comply with this policy, team members will do so within a **two-year period** after data have been collected for their particular research project.

However, this will not apply to data deemed sensitive by researchers or their Research Ethics Board (an example might include qualitative data in which research participants describe difficult past housing experiences).



Preservation: Resources

Data Repositories:

- DataCite Repository Finder tool
<https://repositoryfinder.datacite.org/>
- McMaster Dataverse
<https://dataverse.scholarsportal.info/dataverse/mcmaster>
- FRDR <https://www.frdr-dfdr.ca/repo/>
- Data Repository Guidance from *Nature Scientific Data*
<https://www.nature.com/sdata/policies/repositories>

Sharing & Re-use

- What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).
- When will the data be shared?
- Have you considered what type of end-user license to include with your data?
- What steps will be taken to help the research community know that your data exists?
- Are there any methodological or other considerations that preclude data sharing?

Andreas Kay <https://flic.kr/p/ZjmrWN>

Sharing & Re-use

The analyzed, de-identified data set or datasets will be put under **mediated access** in the Scholars Portal Cape Breton University **Dataverse**. Users will be required to request access to the data for reuse.

Requests will be **evaluated by the PI** and/or a backup member identified on the research team. Terms of access and use will be determined by the PI in consultation with the research team to ensure appropriate use of the data.

Data deposited in Dataverse will be assigned a **Digital Object Identifier (DOI)**, a unique and persistent code that can be used by others to locate and access these data. **Metadata is harvested by the FRDR**, a Canada wide research repository, where data can be discovered, and then shared, at a national level. We will also link our dataset to the **publications** arising from this study.



Sharing & Re-use: Resources

Data Anonymization:

- Portage Network De-Identification Guidance
<https://zenodo.org/record/4270551>

Data Licenses:

- DCC How to License Research Data
<https://www.dcc.ac.uk/guidance/how-guides/license-research-data>

Responsibilities & Resources

- Who will be responsible for managing this project's data during and after the project?
- What are the data-related roles and responsibilities for other team members?
- How will responsibilities for managing data be handled if substantive changes happen in the personnel overseeing the project's data, including a change of Principal Investigator?
- What resources will you require to implement your data management plan? What do you estimate the overall cost for data management to be?

Responsibilities & Resources

The **project lead** is responsible for ensuring team members follow this data management plan. University-based team members are responsible for informing their student researchers/HQPs of this plan.

A **backup member** of the research team will be identified in case the project lead can no longer complete their duties.

Cape Breton University Library offers Dataverse services for the university at **no cost** to researchers. Storage of data in external drives, and other related expenses, could cost approximately **\$200.00-\$300.00**.



Responsibilities & Resources: Resources

Costing

- UK Data Service Costing tool and checklist
<https://ukdataservice.ac.uk/media/622368/costingtool.pdf>
- University of Utrecht Costs of data management estimator
<https://www.uu.nl/en/research/research-data-management/guides/costs-of-data-management>

Roles & Responsibilities

- DataOne Best Practice: Define roles and assign responsibilities for RDM <https://dataoneorg.github.io/Education/bestpractices/define-roles-and>



Ethics & Legal Compliance

If your project involves data from Indigenous communities, DMPs must be co-developed with them in accordance with RDM principles or DMP formats that they accept.

- If your research project includes sensitive data, how will you ensure that it is securely managed and accessible only to approved members of the project?
- If applicable, what strategies will you undertake to address secondary uses of sensitive data?
- What ethical, legal, and commercial constraints (if any) are the data subject to?

Ethics & Legal Compliance

No sensitive data will be shared. Any sensitive data will be stored on secure servers for 5 years.

Research has been approved by the Research Ethics committees at the various institutions involved in the project. Participants are also required to sign the informed consent agreement. By mediating data requests and determining their own terms of access, researchers maintain their rights to the intellectual property.



Ethics & Legal Compliance: Resources

Indigenous Data Principles:

- First Nations OCAP Principles <https://fnigc.ca/ocap-training/>
- CARE Principles for Indigenous Data Governance <https://www.gida-global.org/care>

Securing sensitive data:

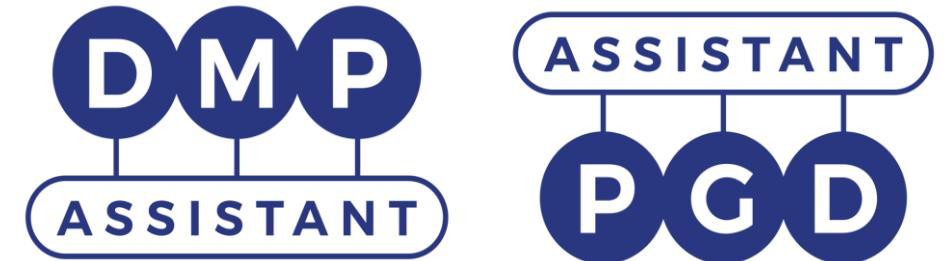
- Portage Sensitive Data Toolkit
 - Glossary of Terms for Sensitive Data <https://zenodo.org/record/4088946>
 - Human Participant Research Data Risk Matrix <https://zenodo.org/record/4088954>
 - Research Data Management Language for Informed Consent <https://zenodo.org/record/4107178>
- McMaster RDM page 'Secure' <https://rdm.mcmaster.ca/secure>

General Best Practices & Resources

- Be specific in your answers
- Review your DMP regularly and revise it when things change
- Follow exemplars
 - Alliance DMP Exemplars and templates
<https://alliancecan.ca/en/services/research-data-management/learning-and-training/training-resources#heading-dmp-exemplars>
- Contact us if you need help – rdm@mcmaster.ca
- McMaster RDM Website <https://rdm.mcmaster.ca/>

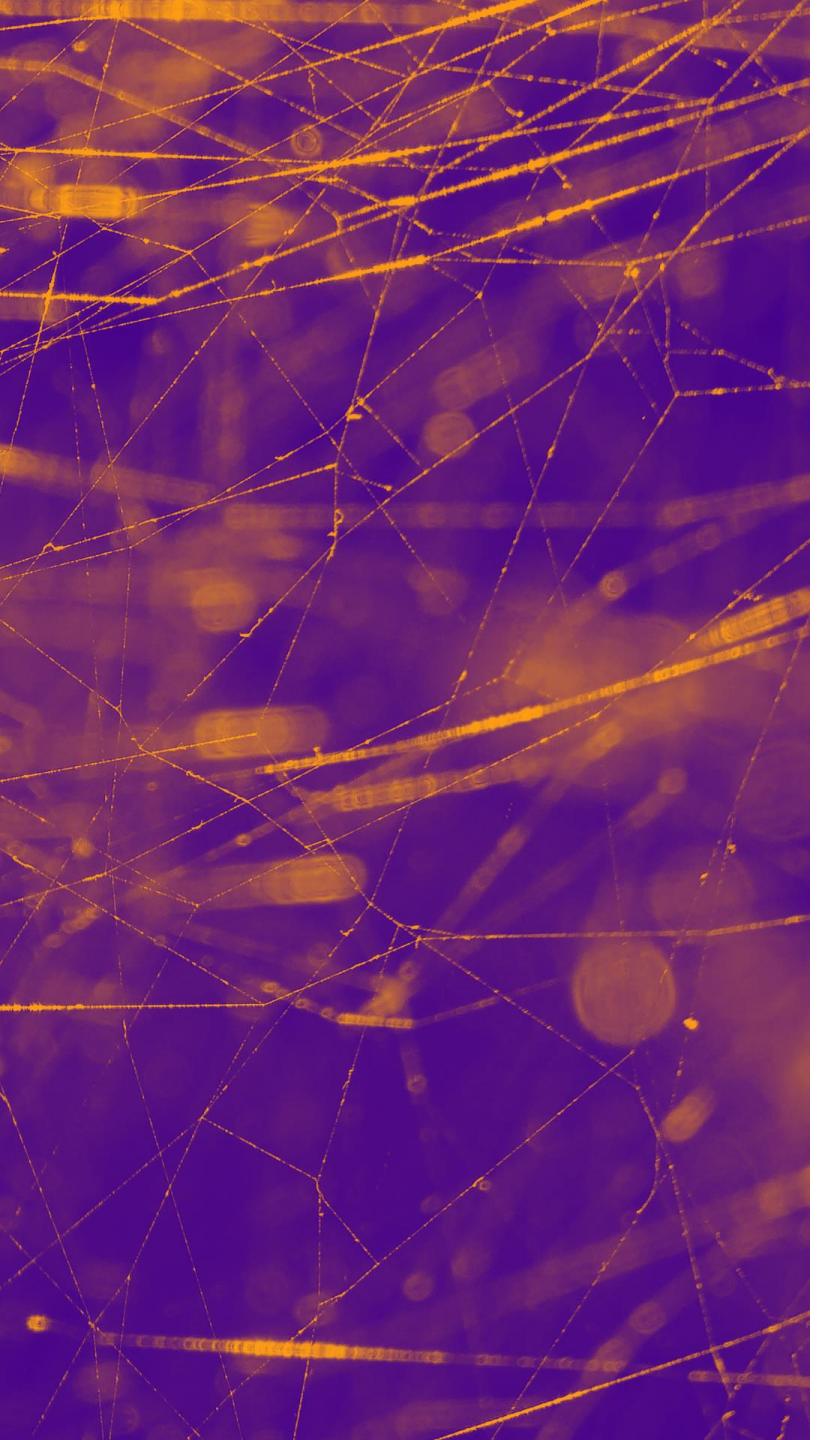
DMP Assistant

- 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



A screenshot of the DMP Assistant web application. At the top, there's a navigation bar with links for 'My Dashboard', 'Create plans', 'Reference', 'Help', 'Language' (set to English), and 'Isaac Pratt'. Below the navigation is a section titled 'My plan (Portage Template)'. This section includes tabs for 'Project Details', 'Plan overview' (which is selected), 'Write Plan', 'Share', 'Request feedback', and 'Download'. There's also a link to 'expand all | collapse all'. The main content area shows a list of categories with counts: 'Data Collection (0 / 3)', 'Documentation and Metadata (0 / 3)', 'Storage and Backup (0 / 3)', 'Preservation (0 / 2)', 'Sharing and Reuse (0 / 3)', 'Responsibilities and Resources (0 / 3)', and 'Ethics and Legal Compliance (0 / 3)'. Each category has a '+' sign to its right.

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



RDM Community of Practice

- Monthly meetings of people interested in RDM at McMaster – **Thurs. Nov. 24 – 11 AM!**
- Guest presentation by Dr. Lawrence Grierson - Family Medicine at McMaster
- Exploring RDM in Medical Education Research.
- Connect with other researchers practicing RDM across the university!
- <https://u.mcmaster.ca/rdm-community>



December 7 | 10:30-11:30am
Virtual Workshop

The Tri-Agency Research Data Management Policy

u.mcmaster.ca/scds-events

SCDS

Library

McMaster
University



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

Depositing & Sharing Data Online with McMaster Dataverse

u.mcmaster.ca/scds-events

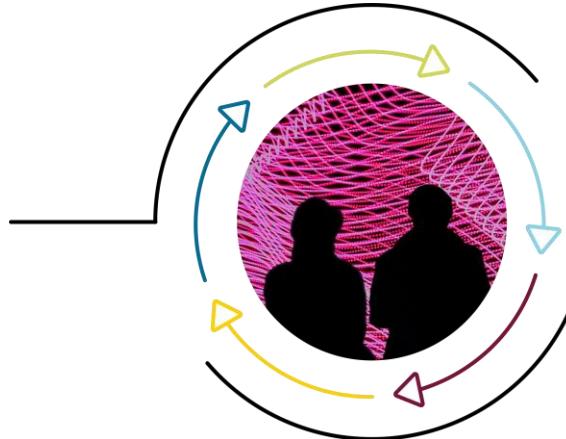


SCDS

Library

McMaster
University





Research Data Management

Services

McMaster RDM webpage:

rdm.mcmaster.ca

Contact RDM services at:

rdm@mcmaster.ca

Upcoming RDM webinars:

rdm.mcmaster.ca/events

Recorded RDM webinars:

u.mcmaster.ca/learn-rdm

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

