



Building a Data Management Plan for your research project.

Isaac Pratt, PhD

May 6th, 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.

SESSION RECORDING

- 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.

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/

HELLO! A bit about me:

- I am a Research Data Management Specialist
- My background is in Biological Anthropology, Medical Imaging, and Human Anatomy
- I have a PhD in Anatomy & Cell Biology from the University of Saskatchewan

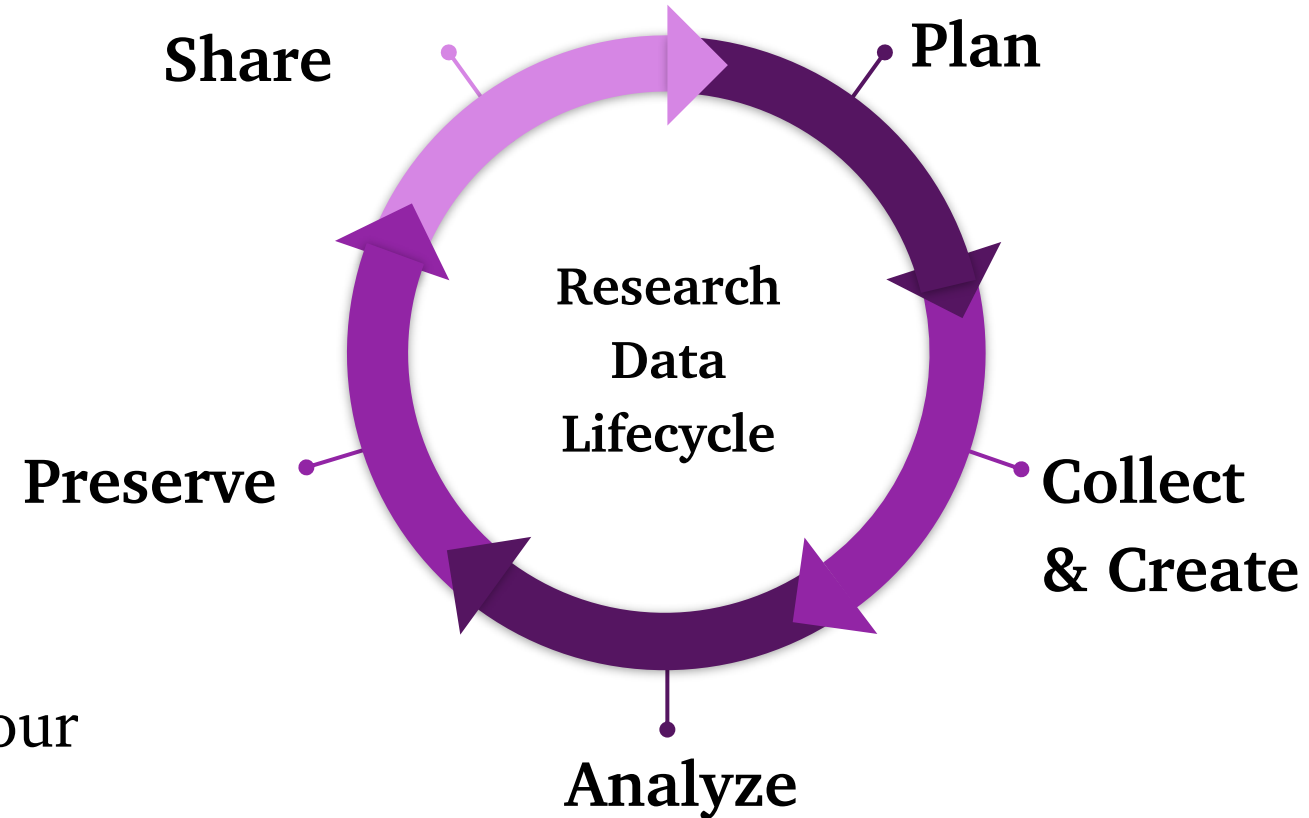
OUTLINE FOR TODAY

- What is RDM?
- RDM Policy
- Data Management Plans
 - Why are they important?
 - What goes in one?
 - What makes a good plan?
- Portage 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/>



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
- **Depositing and Publishing data** increases the visibility of research



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.**

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.

“One who does not plan long ahead will find trouble at his door” - Confucius

Consider the most common approach to data management:

- Data is stored on laptop or desktop hard drives and backed up to a collection of miscellaneous external hard drives accumulated over the years.
- Data is not consistently documented
- 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

DATA MANAGEMENT PLANNING

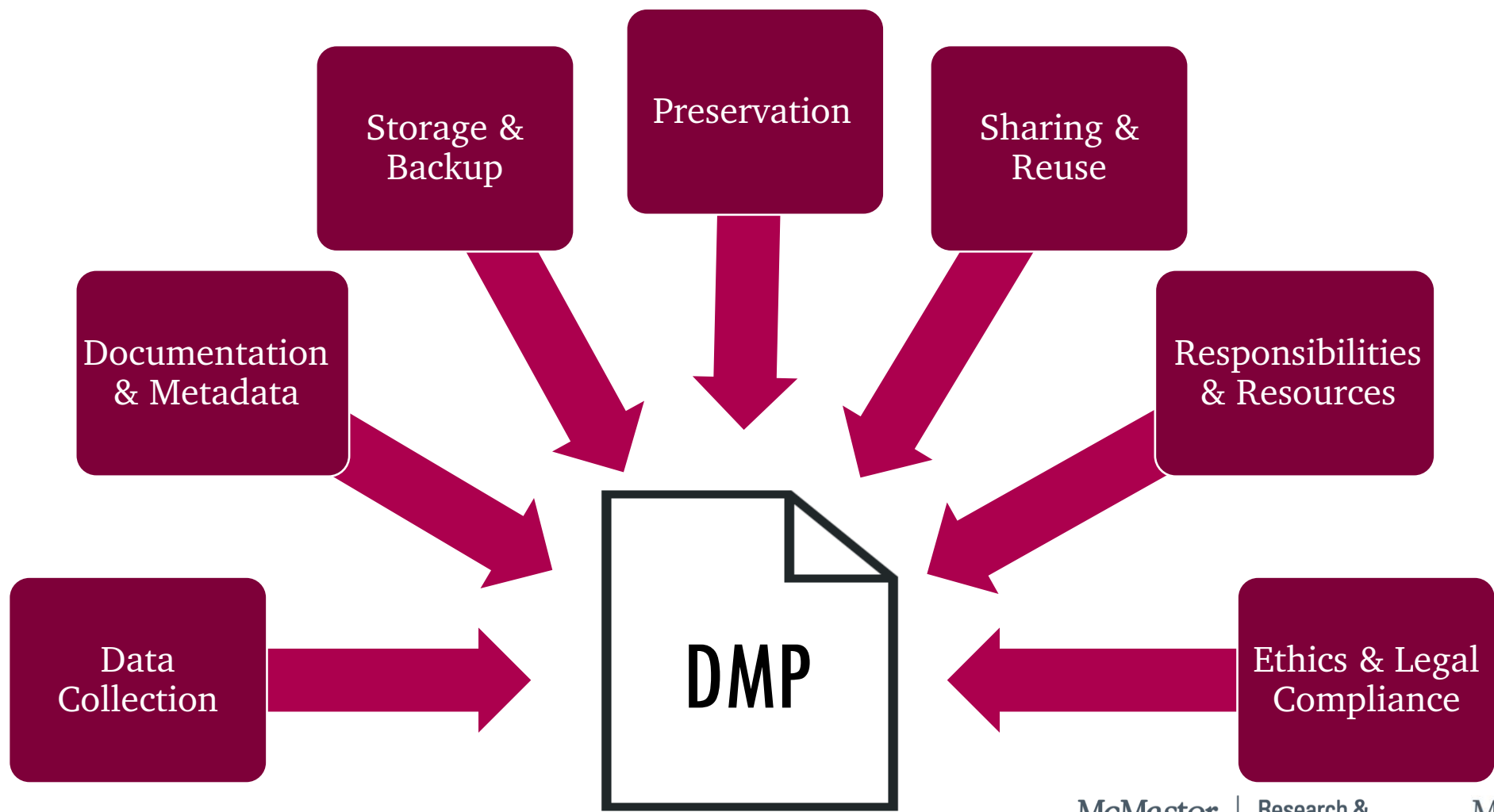
A **Data Management Plan (DMP)** is a living document describing your plan for how you will create, store, organize, document, secure, preserve, and share your research data.

- Creating your own DMP is straightforward using web tools such as the [Portage DMP Assistant](#)
- 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.
- DMPs guide researchers in articulating their plans for managing data; they do not compel researchers to manage data differently.

WHY CREATE A DMP?

- 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



DMP EXEMPLAR: Data Management Plan for People, Places, Policies and Prospects

- Portage Network has published a number of DMP exemplars available here: <https://portagenetwork.ca/tools-and-resources/training-resources/>
- We will 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 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 different file formats
<https://datacurationnetwork.org/outputs/data-curation-primers/>
- McMaster Library webpage on file naming and organization
 - <https://library.mcmaster.ca/services/rdm#tab-organizing-research-files-and-folders>

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

- Cornell ReadMe template

<https://cornell.app.box.com/v/ReadmeTemplate>

Metadata

- Portage 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>

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:

- 3-2-1 Rule for Backup:
 - 3 copies of your data
 - 2 copies backed up on different storage platforms
 - 1 copy stored offsite
- LOCKSS – “Lots Of Copies Keeps Stuff Safe”

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:

- Portage Network Portage Repository Options Guide
https://portagenetwork.ca/wp-content/uploads/2019/06/Repository-Options-Guide_EN.pdf
- DataCite Repository Finder tool
<https://repositoryfinder.datacite.org/>
- McMaster Dataverse
<https://dataverse.scholarsportal.info/dataverse/mcmaster>
- FRDR <https://www.frdr-dfdr.ca/repo/>

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?

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>

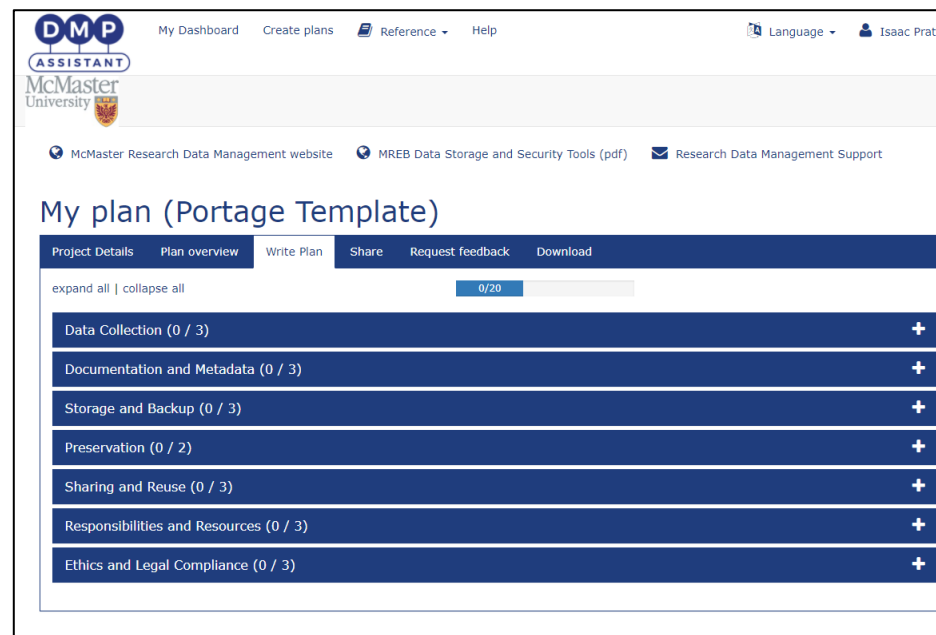
Securing sensitive data:

- Portage Sensitive Data Toolkit <https://portagenetwork.ca/news/new-sensitive-data-tools/>
- McMaster Library 'Secure' RDM page <https://library.mcmaster.ca/services/rdm#tab-secure-your-data>

GENERAL BEST PRACTICES & RESOURCES

- Be specific in your answers
- Review your DMP regularly and revise it when things change
- Follow exemplars
 - Portage DMP Exemplars and templates
<https://portagenetwork.ca/tools-and-resources/training-resources/>
- Contact us if you need help – rdm@mcmaster.ca
- McMaster Library RDM Website
<https://library.mcmaster.ca/services/rdm>

- 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/>

THANK YOU!

For more information:

Visit: library.mcmaster.ca/services/rdm

Contact me at: rdm@mcmaster.ca

RDM
@McMaster