



Research Data Management: From Plan to Preservation!

Presented by staff of the Digital Research Alliance of Canada

Today's Presenters

Nichole DeMichelis - RDM Basics

Robyn Nicholson - Data Management Planning

Tamanna Moharana - Active Data Management

What are Research Data?

The **Tri-Agency** definition:

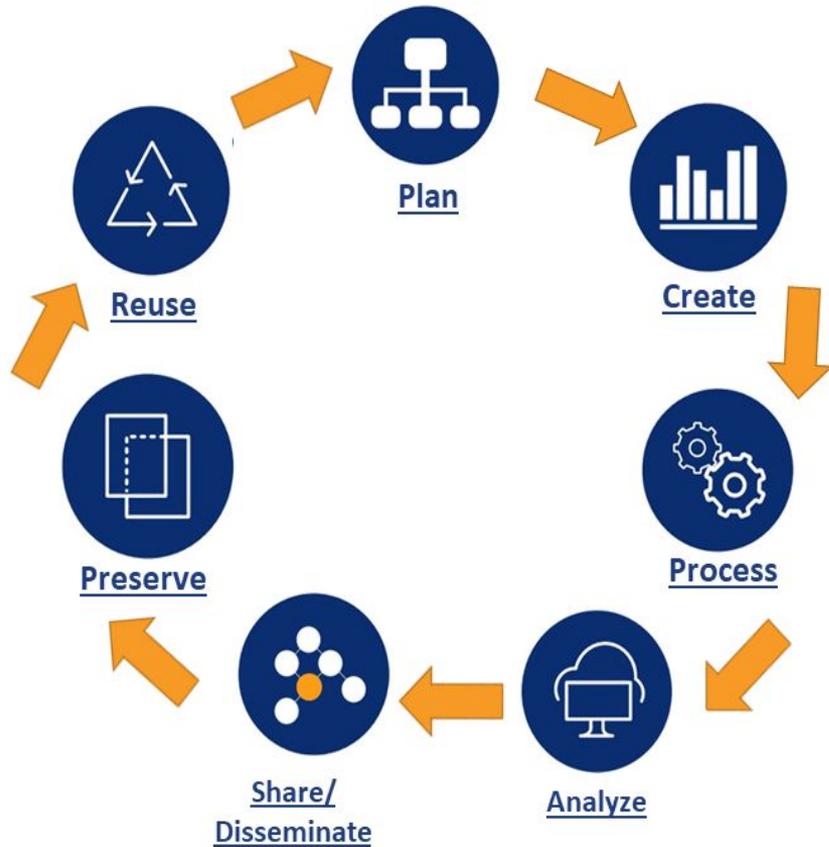
“data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or creative practice, and that are used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results.”

What is research data management (RDM)?

The range of processes and procedures “applied through the lifecycle of a research project to guide the collection, documentation, storage, sharing and preservation of research data”

(“1d: What is research data management?”, Frequently Asked Questions: Tri-Agency Research Data Management Policy, last accessed January 6, 2022 at https://www.science.gc.ca/eic/site/063.nsf/eng/h_97609.html#1d)

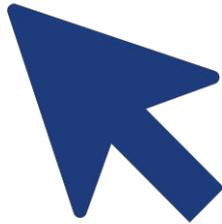
The Research Data Life Cycle



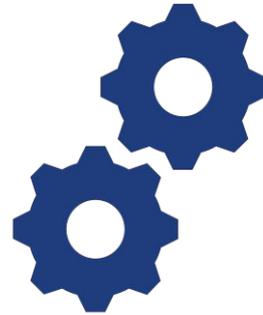
The FAIR Principles



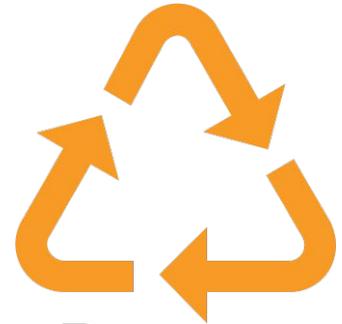
Findable



Accessible



Interoperable



Reusable



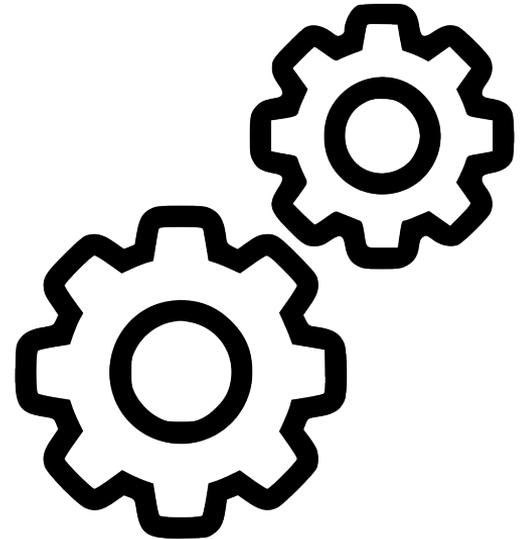
Why RDM? The Big Picture View

- Moves towards open/transparent science
- Recognition of research data as an asset
- Reducing costs of research/Responsible use of public funds
- Reproducibility

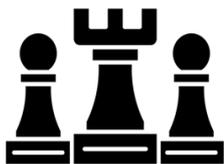


Why RDM? The Day-to-Day View

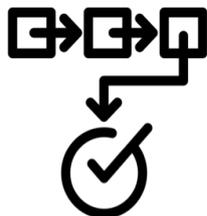
- Making everyday research work more efficient and effective
- Ethical, privacy, and security concerns
- Increasing research impact
- Funder and publisher requirements



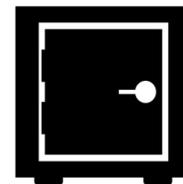
The Tri-Agency RDM Policy



**Institutional
Strategy**



**Data
Management
Plans**



**Data
Deposit**



Data Management Plans

What is a DMP and how can it help you?



Data Management Plans: What, Why, & When?

What is a Data Management Plan (DMP)?

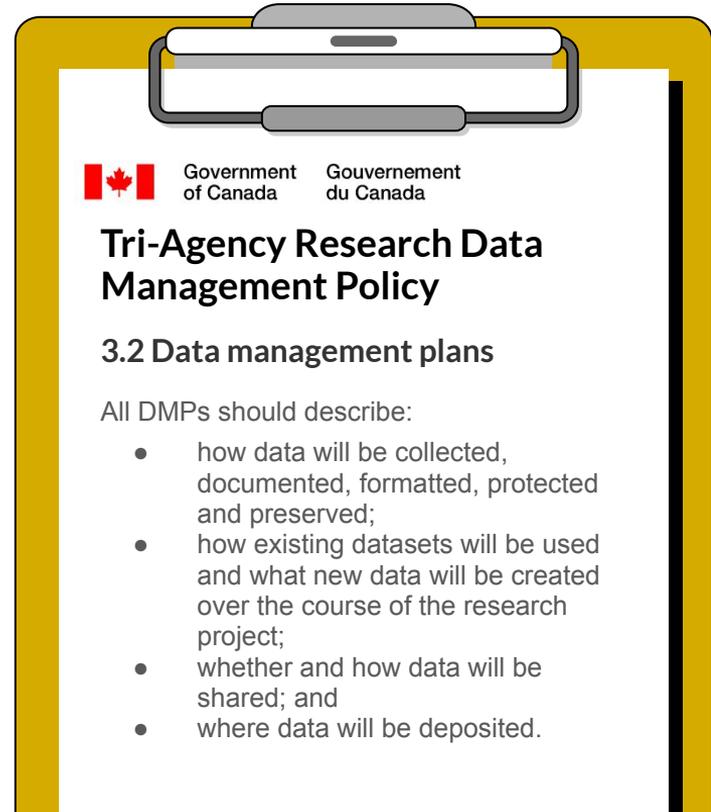
- A formal document which clearly articulates the strategies and tools you will implement to effectively manage your data
- Speaks to the management of data both **during** the active phases of your research and **after** the completion of the research project
- A **“living”** document that can be modified throughout your project to reflect any changes that have occurred

Objective:

To address issues related to data management prior to starting your research project!

Why are DMPs important?

- Because it is **good practice**.
- Because it is **practical**.
- Because it *is* and/or *will be* **required**.



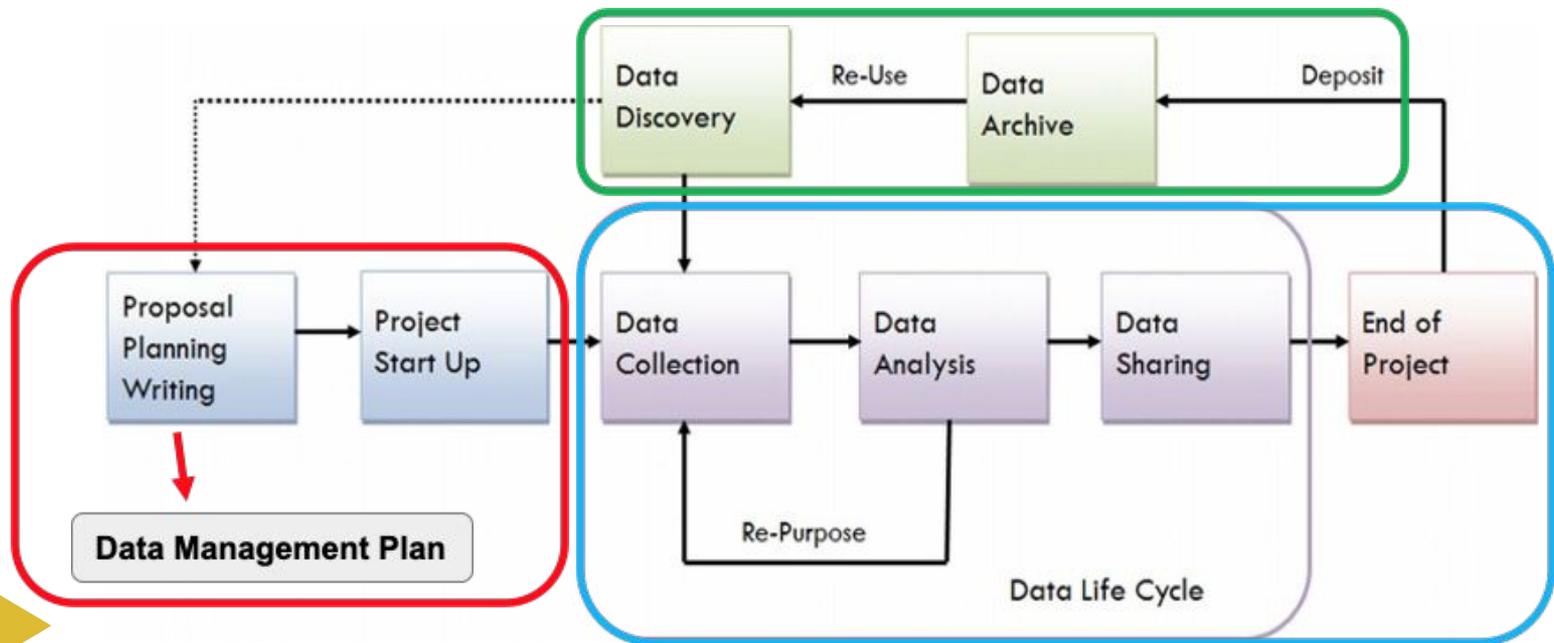
Why are DMPs important?

A DMP is important to the research process because it can help you to:

- **Set out consistent strategies** prior to starting research for how data will be managed throughout its entire lifecycle
- **Identify strengths & weaknesses** in current practices and make decisions on how to adopt effective data management practices
- **Prepare data** for future reuse, preservation and sharing
- **Plan for** and **reduce overall cost** of research by improving project efficiencies and data management practices.

DMPs & the Research Lifecycle

RDM - managing data throughout all phases of the research lifecycle through **planning**, **active research**, and **beyond!**

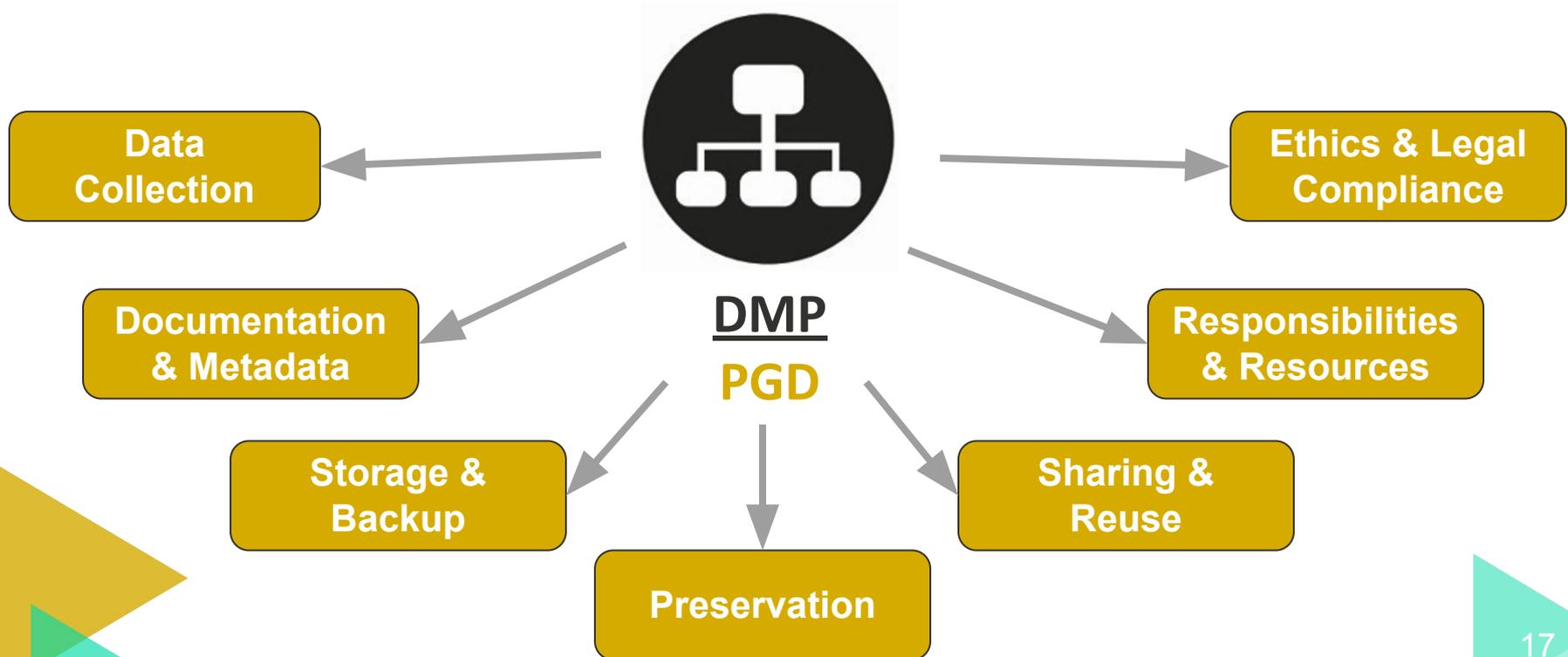


Source: Research Data Services + Sciences (2021). Research Data Lifecycle. In *Research Data Management*. University of Virginia.
<https://guides.lib.virginia.edu/c.php?g=515290&p=3522215>.

DMP Contents

DMP Contents

A DMP provides information across key research lifecycle categories:



Include descriptions of **how you will collect data**, including **from where** and in what **format(s)**

Provide an **estimate of the amount of data** you will collect (e.g., *MBs/GBs/TBs*).

Describe any **software** and/or **platforms** that will be used for data collection

Data Collection

Clearly explain **how** you will **both store** and **transfer** data

Explain **how you will organize your data**, including details relating both to **file naming** and **versioning**

Documentation & Metadata

Describe what information will be needed **for others to understand or reuse your data**

Describe how you will **consistently capture documentation** throughout the project

Choose a **metadata standard** suited to your *discipline* and/or chosen *data repository*, or **provide rationale** for creating your own.

Storage & Backup

Provide an **estimate of storage space needed** during the active phases of your research:
Remember to take into account file versioning, backups, and data growth!

State a **data backup schedule**, automatic being most ideal

Describe how collaborators or research team will **access, modify, contribute**, and **work with** your data.

Preservation

Not all data that you create necessarily *needs* to be preserved: Consider such things as the **value of your data, funding requirements**, etc., and decide which, if any, should be preserved.

Consider **optimal file formats** for supporting **long-term preservation**: Optimally preserved data are **easily accessible** and used by anyone, *without* requiring proprietary software.

Consult with **experts in the Libraries** as needed!

Consider the **appropriate sharing** of your data, including any **funding** or **confidentiality requirements**

Explain what uses can be made of your data through licenses like **Creative Commons**

Consult with colleagues or librarians to choose an **appropriate data repository**, or search **re3data.org**

Sharing & Reuse

Choose a repository that assigns **permanent identifiers** to datasets (e.g., **DOI**) to enhance *discoverability, accessibility, and citability.*

If applicable, describe how you will **ensure file integrity, anonymization** and **de-identification.**

Responsibilities & Resources

Identify **data stewardship roles** and **responsibilities** of project members and other organizations during and after the project

Estimate and describe any **required resources** and **costs** for **data management** and **long-term access** to your data.

Ethics & Legal Compliance

Describe if there are any **legal, ethical, and intellectual property issues** when managing and sharing your data.

Explain how you will comply with any **applicable privacy legislation**, funding and institutional **requirements**.

Describe how you will ensure data is **securely managed** *after* the project is completed, in accordance with any **ethical obligations**, including management of **sensitive data**

Begin by providing a **description** of your research project, its **focus**, and **purpose**

Provide **clarification** for any **acronyms** used

Avoid discipline specific **jargon**: Your DMP should be **easily understood** by anyone!

General Guidelines for DMPs

Avoid leaving sections or questions **blank**

Your DMP is a **living document**: Update it as needed!

Provide **rationale for decisions**: Help others understand *why* you have made a decision!

General Guidelines for DMPs



Length will depend on **complexity, discipline, institutional** and/or **funder requirements**:
Generally **~1-5 pages**



You can **reuse & recycle** DMP content for similar projects and for **funding** or **ethics board applications!**

DMP Supports & Resources

For more information on DMPs:



DATA MANAGEMENT PLANS

This brief guide provides basic information about data management plans (DMPs). More resources for DMPs can be found on the [Portage website](#), including a [Brief Guide for creating an effective DMP](#).

WHAT IS A DATA MANAGEMENT PLAN (DMP)?

A DMP is a formal document that details the strategies and tools you will implement to effectively manage your data both *during* your research project and *after* its completion.

WHY CREATE A DMP?

Efficiency - identify both strategies and potential challenges in advance; develop sound data practices for your research team; prepare data for effective use during your project.

Research Quality - ensure reliability and accuracy of data through careful documentation of your data collection, handling and stewardship practices.

Reusability and Impact - Improve discoverability, accessibility, and reusability of your data by planning for sharing in a repository; increase the potential impact of your research!

Compliance - Satisfy DMP requirements that may be set forth by specific granting agencies or even your own institution.

COMPONENTS OF A DMP

- **Data collection:** data types, file formats, naming and version control
- **Documentation:** ensure data can be read and interpreted
- **Data storage and backup throughout the research**
- **Data preservation strategy for long-term access**
- **Provisions for sharing and reuse**
- **Data management responsibilities and resources**
- **Ethical and legal compliance**



CREATE AN EFFECTIVE DATA MANAGEMENT PLAN

This brief guide presents a general framework for creating an effective data management plan (DMP) to help you plan and organize your research and to meet research funder requirements.

To prepare your DMP, visit the [Portage DMP Assistant tool](#).

GENERAL GUIDELINES

- Begin by providing a description of your research project, its focus, and purpose.
- Avoid the extensive use of discipline specific jargon - your DMP should be easily understood by anyone.
- Provide clarification for any acronyms used.
- Do not leave sections or questions blank.
- Provide rationale for decisions made - help others understand why you have made a decision.
- Your DMP is a living document - update it as needed!

DATA COLLECTION

- Include descriptions of how you will collect data, including from where and in what format(s).
- Provide an estimate of the amount of data you will collect (e.g., MBs/GBs/TBs).
- Describe any software and/or platforms that will be used for data collection.
- Clearly explain how you will both store and transfer data.
- Explain how you will organize your data, including details relating both to file naming and versioning.

DOCUMENTATION AND METADATA

- Describe what information will be needed for others to understand or reuse your data.
- Describe how you will consistently capture documentation throughout the project.
- Choose a metadata standard suited to your discipline and/or chosen data repository or provide rationale for creating your own.



Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
Portage Network | portage@carl-abrc.ca | portagenetwork.ca



DATA MANAGEMENT PLANS

What is a data management plan (DMP)?

A DMP is a formal document that details the strategies and tools you will implement to effectively manage your data during the active phase of your research, and the mechanisms you will use for preserving and appropriately sharing your data at the end of the project. A DMP is a "living" document that can be modified throughout your project to reflect any changes that have occurred. More resources for DMPs can be found on the [Portage website](#), including a [Brief Guide for creating an effective DMP](#).

Why create a DMP?

A DMP helps you:

- Meet grant application requirements and/or adhere to institutional data mandates.
- Make it easier for all team members to document, understand, find, and use the data.
- Plan the resources, tools, and expertise needed for data management.
- Identify challenges for storing, handling, and managing the types and volume of data.
- Ensure reliability, authenticity, accuracy, and reproducibility of your data.
- Have a detailed account of your data collection, handling, and stewardship practices.
- Plan how to make your data FAIR (findable, accessible, interoperable, and reusable) to maximize the research potential and impact of your data.

Standard Components of a DMP

Data collection:

- Describe the data that you will be collecting, including the type, format, and volume.
- Describe *how* you will be collecting your data.
- Establish standards for naming and organizing data files, folders, and versions.

Documentation and metadata (description of data):

- Describe how you will ensure that your data are understandable, interpretable, and usable both by current and future researchers.
- Provide descriptive information for your data to be discoverable once deposited at the end of your project.

DMP Assistant



- A **national, open, bilingual** data management planning (DMP) tool to help researchers better manage their data throughout the lifespan of a project.
- Hosted by the **University of Alberta Library** and supported by the **Alliance**.
- **Develops a DMP** by prompting researchers to answer a number of key data management questions, supported by best-practice guidance and examples.

assistant.portagenetwork.ca



Digital Research
Alliance of Canada

Alliance de recherche
numérique du Canada



UNIVERSITY OF ALBERTA
LIBRARY

DMP Assistant



Researchers can:

- **Create project plans**, using a generic template or an organization-, discipline-, or methodology-specific template;
- **See guidance** appropriate to the particular template they are using and their organization;
- **Collaborate** with multiple researchers on a plan;
- **Connect** to local guidance and support for data management at their academic institutions or research organizations.

assistant.portagenetwork.ca

Welcome to DMP Assistant.

The **DMP Assistant** is a national, online, bilingual data management planning tool developed by the **Portage Network** in collaboration with host institution **University of Alberta** to assist researchers in preparing **data management plans (DMPs)**. This tool is freely available to all researchers, and develops a DMP through a series of key data management questions, supported by best-practice guidance and examples.

DMPs are one of the foundations of good research data management (RDM), an international best practice, and increasingly required by institutions and funders. For example, they are a pillar of the Canadian Tri-Agency Research Data Management Policy.

Getting started:

- [Brief Guide – Data Management Plans](#)
- [Brief Guide – Create an Effective Data Management Plan](#)
- [Primer – Data Management Plan](#)
- [How to Manage Your Data](#)
- [Webinars](#):
 - [Support Your Research with DMP Assistant 2.0](#)
 - [Support Your Research with Data Management Planning](#)

For more resources and training materials spanning the entire research data life cycle, see the [Portage Network Training Resources](#).

The DMP Assistant was adapted from the Digital Curation Centre (DCC)'s [DMPonline](#) tool, and uses the DMP Roadmap codebase developed by DCC and the University of California Curation Center (UC3).

assistant.portagenetwork.ca

[Sign in](#) [Create account](#)

* **First Name**

* **Last Name**

* **Email**

Organisation

My organisation isn't listed.

* **Password**

Show password

Security check

I'm not a robot 
reCAPTCHA
Privacy - Terms

* I accept the terms and conditions

[Create account](#)



DMP Templates

In addition to a generic default DMP template, Alliance RDM also offers **13 discipline- and methodology-specific DMP Templates:**

Humanities

- Arts-Based Research
- History & Humanities

Natural Sciences

- Qualitative Health Science
- Neuroimaging in Neurosciences
- Studying Molecular Interactions
- Water Quality
- Systemic Reviews
- Open Science Workflows

Computing & Technology

- Advanced Research Computing
- Interdisciplinary Health Software/Technology Development

Mixed Methods

- Mixed Methods (Surveys & Qualitative Research)

Research Data Centres (CRDCN)

- Accessing Data from Research Data Centres
- Research Data Centres & External Analysis

DMP Assistant: assistant.portagenetwork.ca

Training Resources: portagenetwork.ca/tools-and-resources/training-resources

DMP Exemplars

Alliance RDM also offers **9 discipline- and methodology-specific DMP Exemplars** providing helpful reference points for writing DMPs:

Digital Humanities

- Belgians and French in the Prairies
- Soundscape Study
- Women's Print History Project (1750-1830)

Digital Humanities & Secondary Data

- Historical Canadian Census Data

Mixed Methods

- Mixed Methods Fictional Exemplar

Natural Sciences

- Echohydrology Research Group
- Computation Reproducibility in High-Performance Computing (HPC)

Social Sciences

- Affordable Rental Housing Study
- Usage of Academic Profile Websites

Training Resources: portagenetwork.ca/tools-and-resources/training-resources



Webinars & Videos

Support Your Research Webinar Series

- [Support Your Research with DMPs!](#)
- [Support Your Research with the DMP Assistant!](#)

DMP Assistant Tutorial Video Series

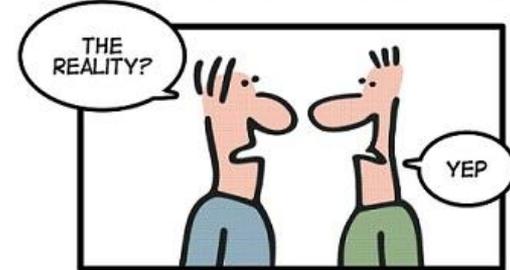
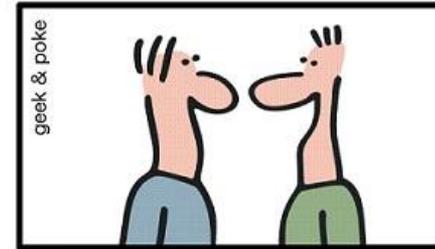
- [Introduction to Data Management Plans \(DMPs\)](#)
- [Introduction to DMP Assistant](#)
- [Managing DMPs with DMP Assistant](#)

Training Resources:

portagenetwork.ca/tools-and-resources/training-resources



Now that we have
planned the work,
Let's Work the Plan!



"PLAN THE WORK, WORK THE PLAN"

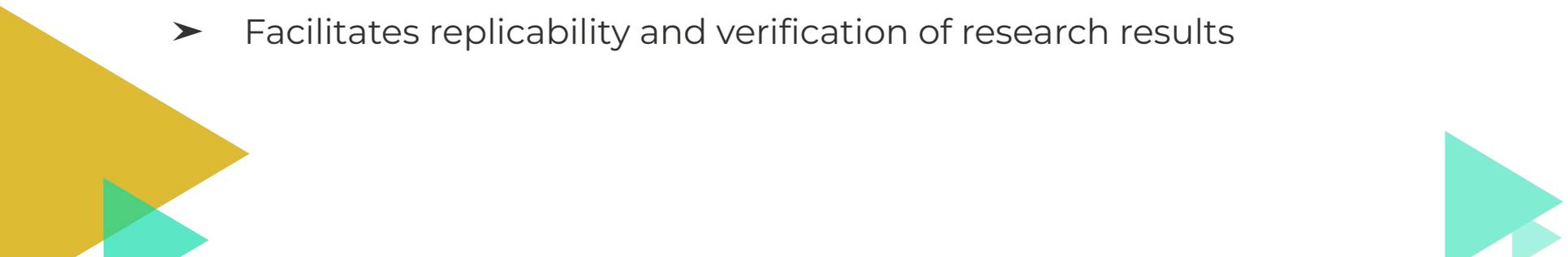




Active Data Management



Benefits for Researchers

- Fulfills Tri-Agency grant requirements
 - Responds to publisher's requirements
 - Ensures research data are accurate, complete, authentic and reliable
 - Improves your research discoverability
 - Saves time and resources in the long run
 - Enhances data security and minimise the risk of data loss
 - Facilitates replicability and verification of research results
- 

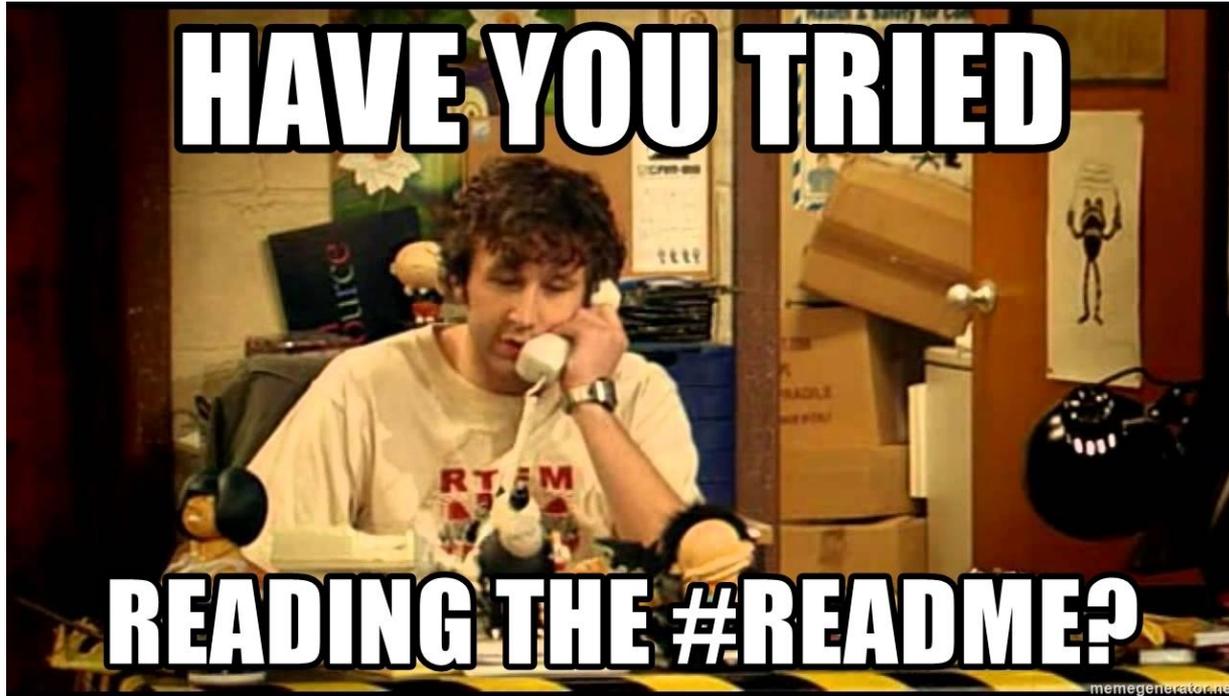
Conversation starters

- How was your data created or gathered?
- What choices did you make along the way?
- Who was involved and what did they do?
- Did you create something new or incorporate existing works and scholarship?
- Are you planning to share your data? If so, where and what considerations are needed to be met to make it shareable?
- Do you have any responsibilities or requirements for your data?
- Where do you see your data in 10 years? In 50 years? What will have changed and what will stay the same?

Beginning with the End in Mind..

- Documentation
- File format transformations
- File inventory
- Arrangement and description
- Evaluate FAIRness
- Metadata
- Data Sharing and Preservation

Documentation is important!



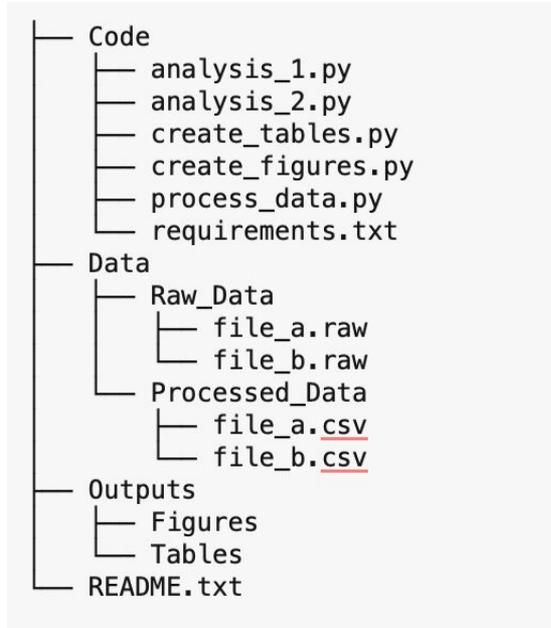
Documentation

- Contextual information about the data (how the data was collected or generated, the goal of the research)
- Description of file naming conventions and the structure of the files, if important
- Record of how the data were modified or processed
- Information about confidentiality and any restrictions placed on secondary use
- Names of labels and variables, information about allowable values and units of measure, codes and classifications, if applicable
- Description of the computing environment required to run any code that has been included (operating system, software packages and dependencies)

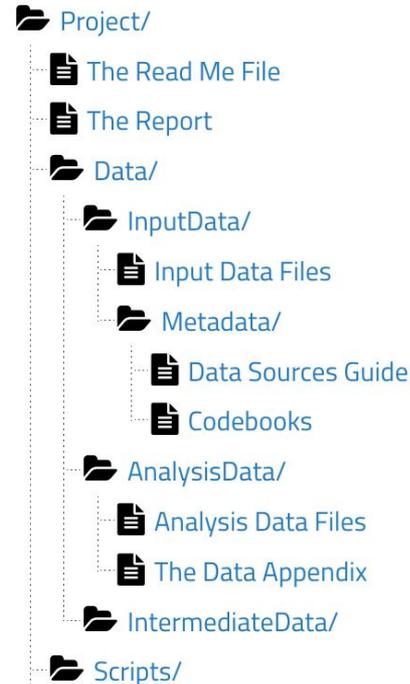
Brigham D. (2020). [Creating a README for Your Dataset](#) (Quick Guide) & [Cornell University "Guide to writing 'readme' style metadata"](#)

Checklist excerpted from Cooper et al. (2021). [Dataverse Curation Guide](#)

File organization



Sample directory structure

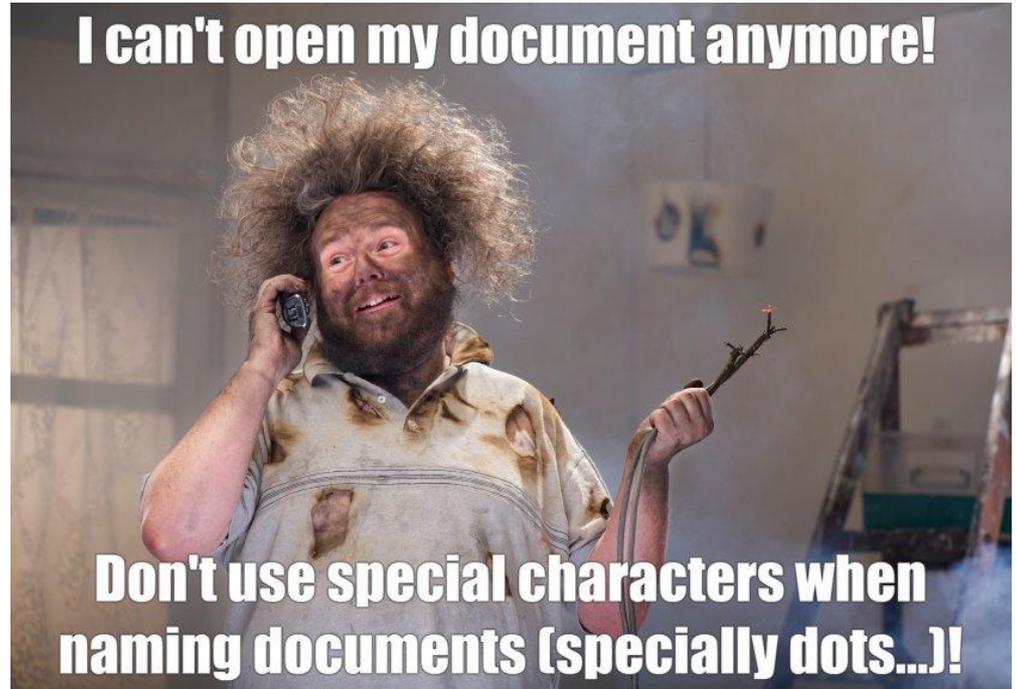


Excerpt from TIER Protocol 4.0

File naming guidelines

Avoid using:

- Spaces (use - or _ instead)
- Special characters (~ ! @ # \$ % ^ & * () ` ; < > ? , [] { } ' " and |)
- Potentially sensitive or restricted information



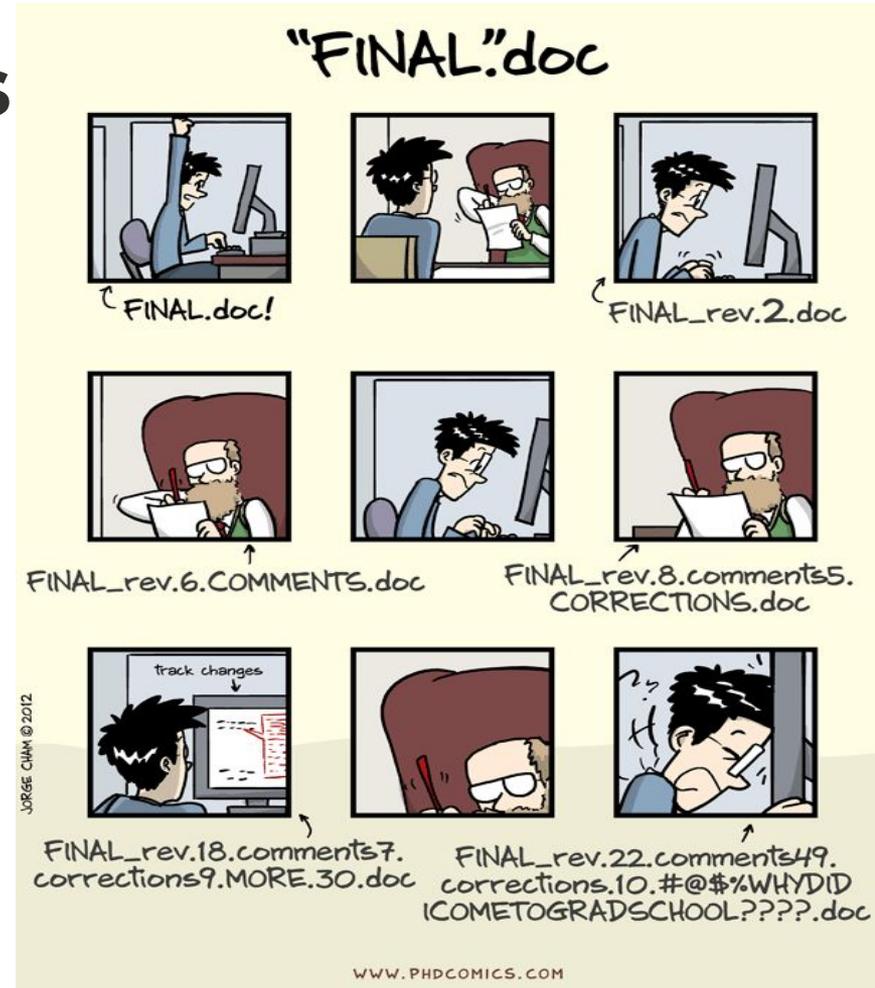
Source: https://twitter.com/f_gelati/status/1247798577205661696

Resource: University of Ottawa Library, File naming and organization of data:

<https://biblio.uottawa.ca/en/services/faculty/research-data-management/file-naming-and-organization-data>

File naming guidelines

- Balance between concise and descriptive.
May contain:
 - Project name, type of data or analysis
 - Date (YYYY-MM-DD format)
 - Version number (v_01, v_02, etc.)
- Lead numbers with a zero (e.g, 001, 002 ... 010, 011 ... 100, 101)
- Try to make file names unique (directory structure may change over time)
- Document naming conventions to help with consistency



What is Metadata?



Key Metadata Elements

➤ Description

- Describe the dataset - not the paper
- Write thinking about search and reuse:
 - Can someone find it?
 - Is there enough context to understand it?

➤ Subjects / keywords

- See if the repository uses controlled vocabulary
- Or use a vocabulary frequently used in your discipline
- Think about keywords you search with

➤ Geospatial - for searching by location

- Useful to include even if your data isn't geospatial - if your data is about a place at all
- Place names, bounding boxes, points
- Can also put place names in keywords

Key Metadata Elements

- Licenses - Creative Commons, etc.
 - All this is about communicating how data can be reused
 - Some repositories will only allow specific licenses, or apply a license automatically
- Resources:
 - Research data: <https://www.dcc.ac.uk/guidance/how-guides/license-research-data>
 - Software: <https://choosealicense.com/>

Metadata Standards

- International Organization for Standardization Geographic information metadata (ISO 19115)
- Climate and Forecast metadata (CF)
- Water Quality Exchange (WQX)

<https://www.epa.gov/waterdata/water-quality-data>

Data Sharing & Preservation

Responsibilities

You will remain:

- The recognized and attributed creator of the dataset
- The contact person for your dataset (or the person you named will remain the contact person)

Things to consider

- Open or non-proprietary file formats
- Open license
- Well documented data files



Where to deposit data?

Repository selection considerations:

- Domain-specific vs. generalist?
- Institutional vs. national or international?
- How much data?

Resources:

- <https://www.re3data.org/>
- <https://fairsharing.org/>

Checklist for Success ...

- Follow best practice guidelines to name and structure files.
- Create a README, codebook, and other documentation necessary to understand your data and interpret it correctly
 - Document data collection instruments, methods, quality control measures, and specialized software needed to view or manipulate data
 - Describe the contents of your directories and/or files
 - Provide attribution to any external data sources
- Write a description for your dataset and choose keywords that will improve discoverability.
- Select a license that respects the constraints of any data you may have reused (for help choosing a creative commons license, see <https://chooser-beta.creativecommons.org/>)

Recommended resources

- Brigham D. (2020). [Creating a README for Your Dataset](#) (Quick Guide) and the
- [Cornell University “Guide to writing ‘readme’ style metadata”](#)
- [Choose the best file formats / Guide concernant les formats recommandés par BAnQ](#)
- [File naming and organization of data](#)
- [Metadata Standards Catalog](#)
- [TIER Protocol 4.0](#)
- Borer et al., 2009, [Some Simple Guidelines for Effective Data Management](#)
- Sensitive Data Toolkit for Researchers ([part 1](#), [part 2](#), [part 3](#))
- [De-identification Guidance](#)
- [Ethical Considerations in the Use of Geospatial Data for Research and Statistics](#)
- [Your local RDM librarian!](#)
- [Data Curation Network \(DCN\) Curation Primers](#)
 - Detail a specific subject, disciplinary area or curation task that can be used as a reference to curate research data.



**Digital Research
Alliance** of Canada

Accelerating Canada's
Research Future.

**Alliance de recherche
numérique** du Canada

Accélérer l'avenir de
la recherche au Canada.

Thank you!

Questions?

FRDR Curation Team

Erin Clary

Nichole DeMichelis

Tamanna Moharana

Y.G. Rancourt

support@frdr-dfdr.ca

Data Management Planning

Robyn Nicholson

robyn.nicholson@alliancecan.ca

DMP Assistant Support

support@portagenetwork.ca

alliancecan.ca

