What is a Data Management Plan and why do I need one?

Presented by:

Dylanne Dearborn
Research Data Management Coordinator,
University of Toronto Libraries

October 2023

UNIVERSITY OF TORONTO

DEFY GRAVITY

Objectives

- Describe why planning and managing research data is important
- Identify common elements often included in data management plans
- Locate resources and supports for developing a data management plan

Agenda

- What is Research Data Management (RDM)
- What is a Data Management Plan (DMP)
- The purpose of a Data Management Plan
- What a Data Management Plan covers
- How to create a Data Management Plan
- Data Management Plan resources

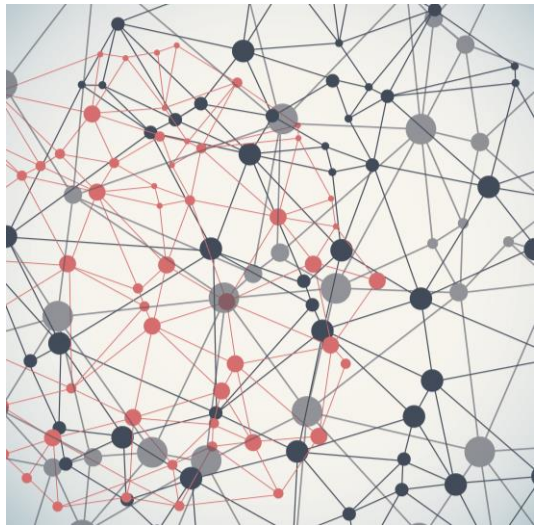
What is Research Data Management

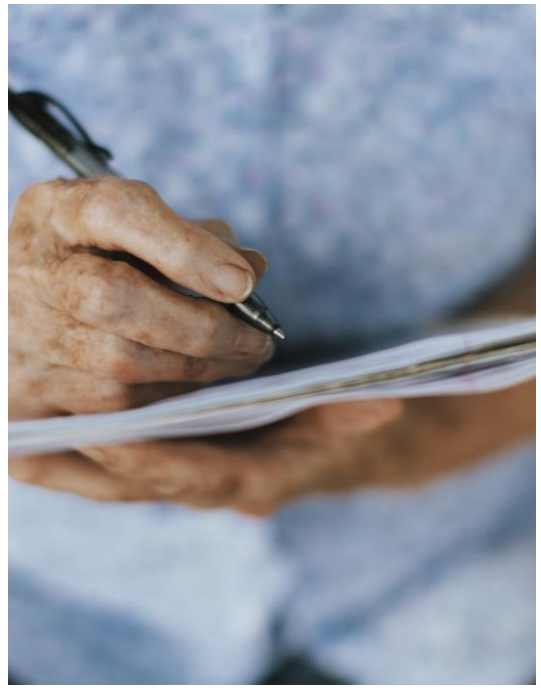
UNIVERSITY OF TORONTO

DEFY GRAVITY

What is Research Data?

- What is collected, observed, or created for the purpose of analysis to produce original research results
- "Research data" will mean different things for different projects
- Includes notes, documentation, scripts, etc.





UNIVERSITY OF TORONTO

What is Research Data Management (RDM)?

- RDM is about handling research data effectively and appropriately throughout a research project, from beginning to end and beyond
- RDM refers to all aspects of planning, creating, storing, sharing and preserving data
- RDM is an essential aspect of conducting responsible research
- Well managed research data means:
  - Others should be able to understand your data
  - Others should be able to reproduce your results

Research Data Management - Planning

Planning means:

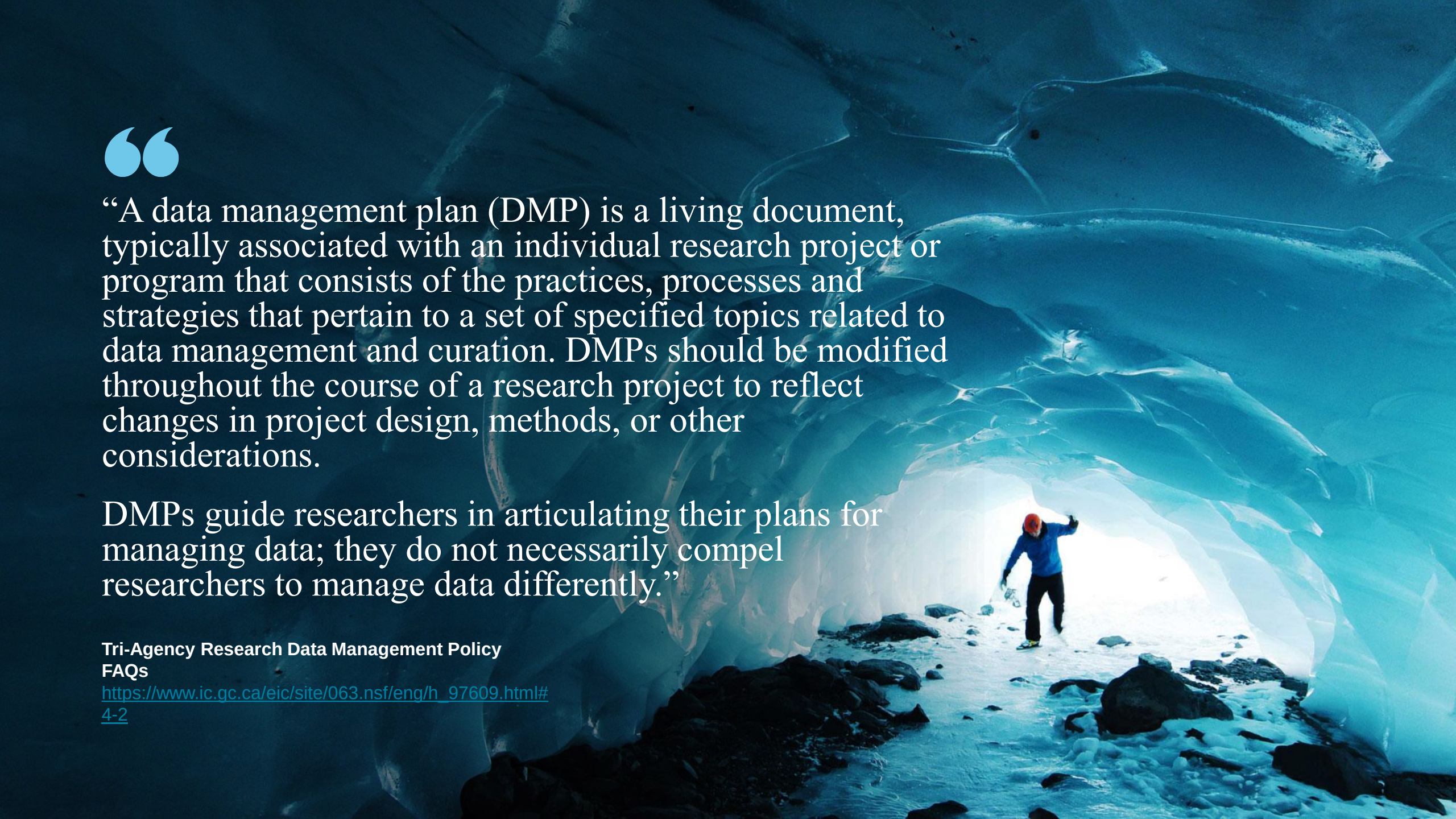
- Preparing for the research project
- Conceptualizing data management actions across project stages
- Developing consistent approaches for working with the data
- Considering what is needed to carry out the strategy

Documenting and formalizing the strategy = **a Data Management Plan (DMP)**

What is a Data Management Plan

UNIVERSITY OF TORONTO

DEFY GRAVITY



“A data management plan (DMP) is a living document, typically associated with an individual research project or program that consists of the practices, processes and strategies that pertain to a set of specified topics related to data management and curation. DMPs should be modified throughout the course of a research project to reflect changes in project design, methods, or other considerations.

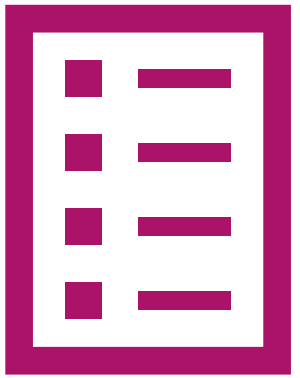
DMPs guide researchers in articulating their plans for managing data; they do not necessarily compel researchers to manage data differently.”

Tri-Agency Research Data Management Policy FAQs
https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97609.html#4-2

What is a Data Management Plan?

A DMP is simply a short, formal document developed at the start of a research project which outlines the plan for all aspects of data management

- Details what you are doing with your data during and after the project



UNIVERSITY OF TORONTO

Typical components of a DMP

- Administrative details
- Data collection
- Documentation and metadata
- Storage and backup
- Retention and preservation
- Sharing and reuse
- Responsibilities and resources
- Ethics and legal compliance

The purpose of a Data Management Plan

UNIVERSITY OF TORONTO

DEFY GRAVITY

How is a DMP beneficial to a project?

- Stay organized
- Improve research quality
- Ensure long term access
- Share your work
- Fulfill emerging requirements

Stay organized

- Find your files and keep track of different versions of your data
- Helps collaborators be on the same page
- Facilitates continuity if project staff leave or new researchers join
- Organize and compile information at the end of a project

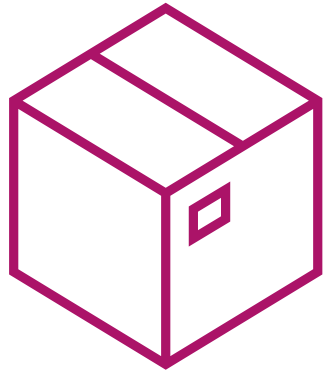
Improve research quality

- Ensures necessary supports are in place and included in budget
- Helps anticipate and mitigate problems, barriers and risk
- Increases project efficiencies, reducing the overall cost and time of research
- Increases reproducibility and transparency
- Guards against misconduct and data falsification

UNIVERSITY OF TORONTO

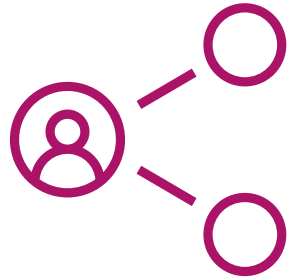
Maintain long-term access

- Helps prepare data packages to retain or deposit into a repository
- Enables preservation actions



Share your work

- Enable the reuse of your data
- Increase visibility and scholarly impact
- Encourage collaboration and interdisciplinary research



Fulfill emerging requirements

- Satisfy granting agency*, partner, and/or publisher requirements

*Further information: Tri-Agency Research Data Management Policy and DMP update

Effective Data Management Plans

- Are detailed and project-specific
- Require thought and considerations for all decisions
- Bear in mind the wider context and consequences of different options
- Have buy-in from all collaborators
- Detail who is responsible for certain actions
- Inventory all data needs across the project
- Include review and evaluation
- Are a living document – update as changes to process are needed (a DMP can be versioned)

What a Data Management Plan covers

UNIVERSITY OF TORONTO

DEFY GRAVITY

Typical components of a DMP

1. Administrative details
2. Data collection
3. Documentation and metadata
4. Storage and backup
5. Retention and preservation
6. Sharing and reuse
7. Responsibilities and resources
8. Ethics and legal compliance

1 | Administrative details

Provides an overview of the research project, including:

- Project details and description
- Who is responsible for the project
- Who is responsible for the data
- Contact information

UNIVERSITY OF TORONTO

2 | Data collection

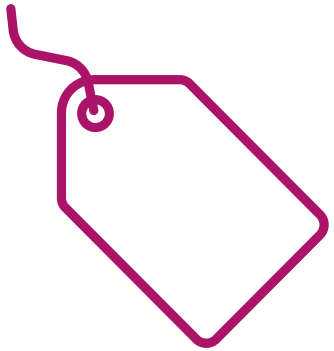
Provides a description of the data, including:

- What data you intend to collect, generate or use
- The different types of data, formats and programs you will be working with
- How you will organize and structure the data

3| Documentation and metadata

Information that allows the data to be understood or reused, including:

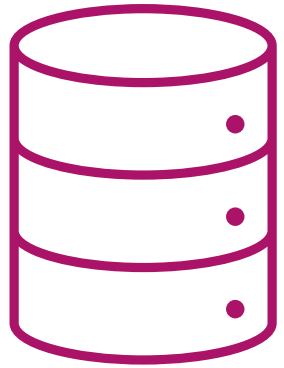
- What documentation will be needed to understand the data
- The metadata standard that will be used to describe the data (if needed)
- How you will capture this information throughout the project



4| Storage and backup

Details the space and procedures for keeping the data during the active phase of research, including:

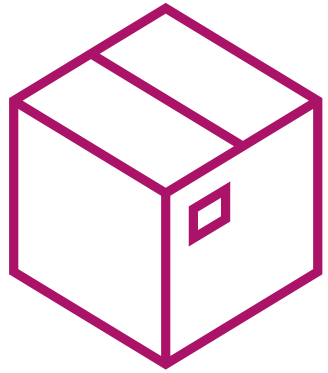
- An estimation of the amount of data being generated
- Where the data will be kept (both working and backup)
- What security measures are needed
- Mechanisms and schedules



5| Retention and preservation

Details what happens to the data at the end of the project, including:

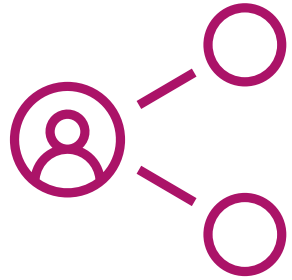
- The files, formats, documentation that will be kept
- How long will data be kept and what will be discarded
- Where will it be kept after the project concludes
- How it will remain accessible over time



6| Sharing and reuse

Details about how you will provide access to external parties, including:

- Any restrictions or reasons data will not be shared
- Strategies to mitigate any issues related to data sharing
- What data will be shared
- Mechanism for sharing
- Conditions of access



7| Responsibilities and resources

Plans the oversight of project provisions, including:

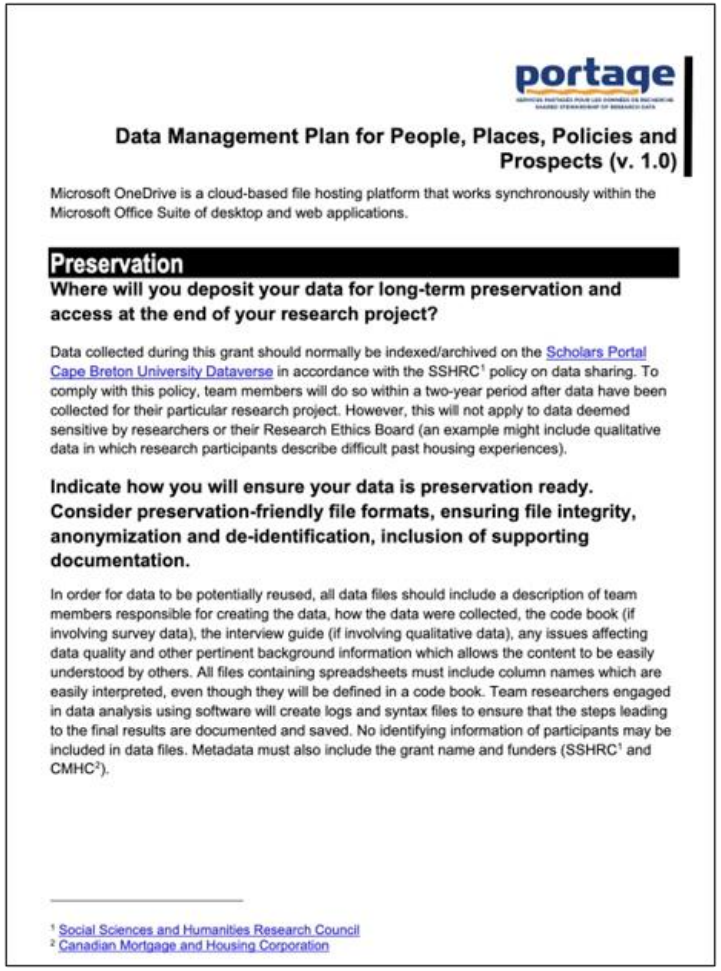
- Roles of team members related to data actions
- The use of available resources
- Budget considerations for personnel or services
- Budget for required tools, software or hardware

8| Ethics and legal compliance

Outlines strategies to ensure data will adhere to applicable requirements, including:

- Privacy legislation and laws, including funding and institutional requirements
- Any legal, ethical, contractual, and intellectual property issues when managing and sharing the data
- Actions that need to be taken across a project to ensure compliance
- Obligations related to the secure management of sensitive data

Example pages of a compiled DMP



portage

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

Microsoft OneDrive is a cloud-based file hosting platform that works synchronously within the Microsoft Office Suite of desktop and web applications.

Preservation

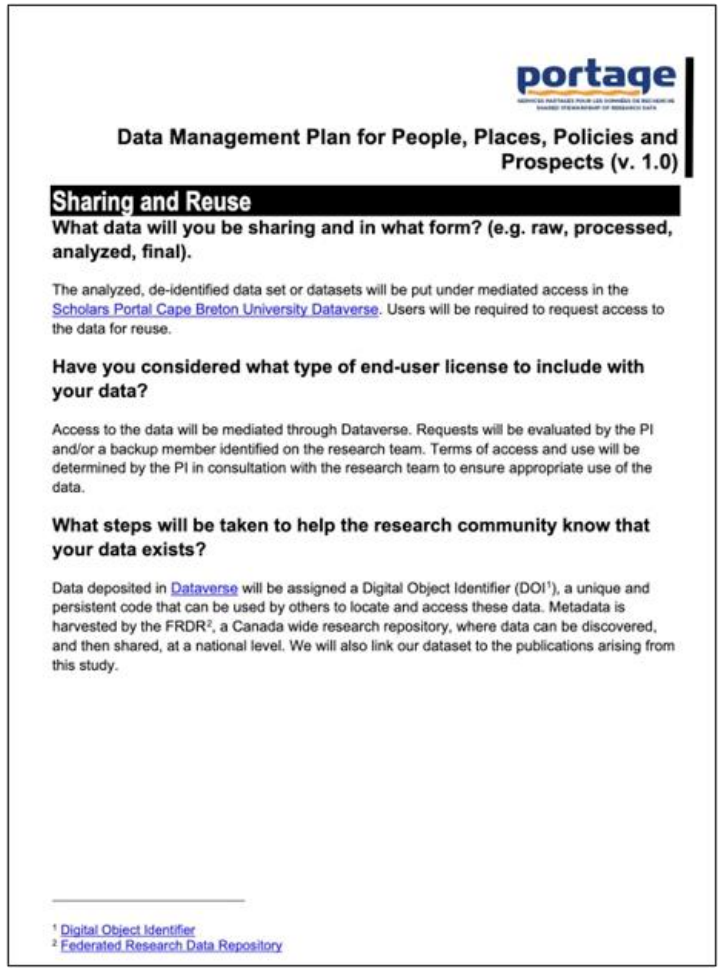
Where will you deposit your data for long-term preservation and access at the end of your research project?

Data collected during this grant should normally be indexed/archived on the Scholars Portal Cape Breton University Dataverse in accordance with the SSHRC[1] 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).

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.

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. 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[1] and CMHC[2]).

[1] Social Sciences and Humanities Research Council
[2] Canadian Mortgage and Housing Corporation



portage

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

Sharing and Reuse

What data will you be sharing and in what form? (e.g. raw, processed, analyzed, final).

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.

Have you considered what type of end-user license to include with your data?

Access to the data will be mediated through Dataverse. 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.

What steps will be taken to help the research community know that your data exists?

Data deposited in Dataverse will be assigned a Digital Object Identifier (DOI[1]), a unique and persistent code that can be used by others to locate and access these data. Metadata is harvested by the FRDR[2], 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.

[1] Digital Object Identifier
[2] Federated Research Data Repository

UNIVERSITY OF TORONTO

Considerations for an effective DMP

- Time consuming up front, but saves time later
- Make it understandable - avoid extended use of acronyms or discipline-specific jargon
- Consider all ethical, cultural, legal or IP considerations
- Consider using existing supports
- Include rationale for decisions
- Do not leave relevant sections or questions blank
- Include sufficient detail for review and evaluation
- Be prepared to implement your plan

How to create a DMP

UNIVERSITY OF TORONTO

DEFY GRAVITY

DMP creation

- Short document to state the RDM plans - consider it a living document
- Can be point form, paragraph, etc.
- There are tools for drafting DMPs:
  - DMPonline (United Kingdom)
  - DMPTool (United States)
  - DMP Assistant (Canada)

What is a DMP template?

A DMP template has a set list of questions that a researcher can use to create a robust plan. Templates can:

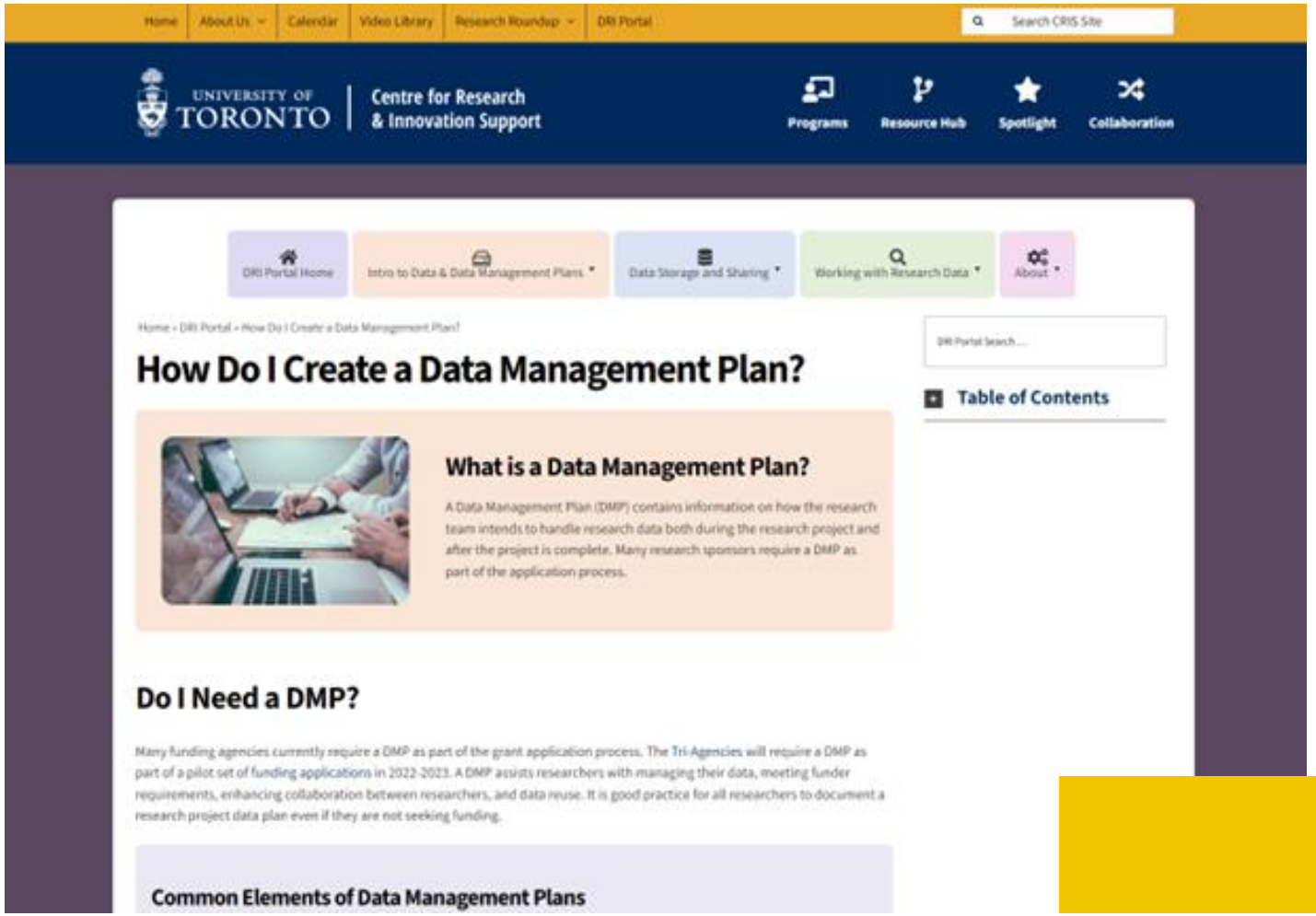
- Come from a funder or institution
  - E.g., UK - Arts and Humanities Research Council (AHRC) DMP template; US - National Science Foundation (NSF) Arctic Data Center: Polar Programs DMP template
- Be for specific disciplines
  - E.g.,. water quality research, art-based research
- Be for specific research methods
  - E.g., systematic reviews, mixed methods (surveys and qualitative research)
- Include tailored guidance to assist in addressing important aspects of the question or providing insight into supports available

Data Management Plan resources

UNIVERSITY OF TORONTO

DEFY GRAVITY

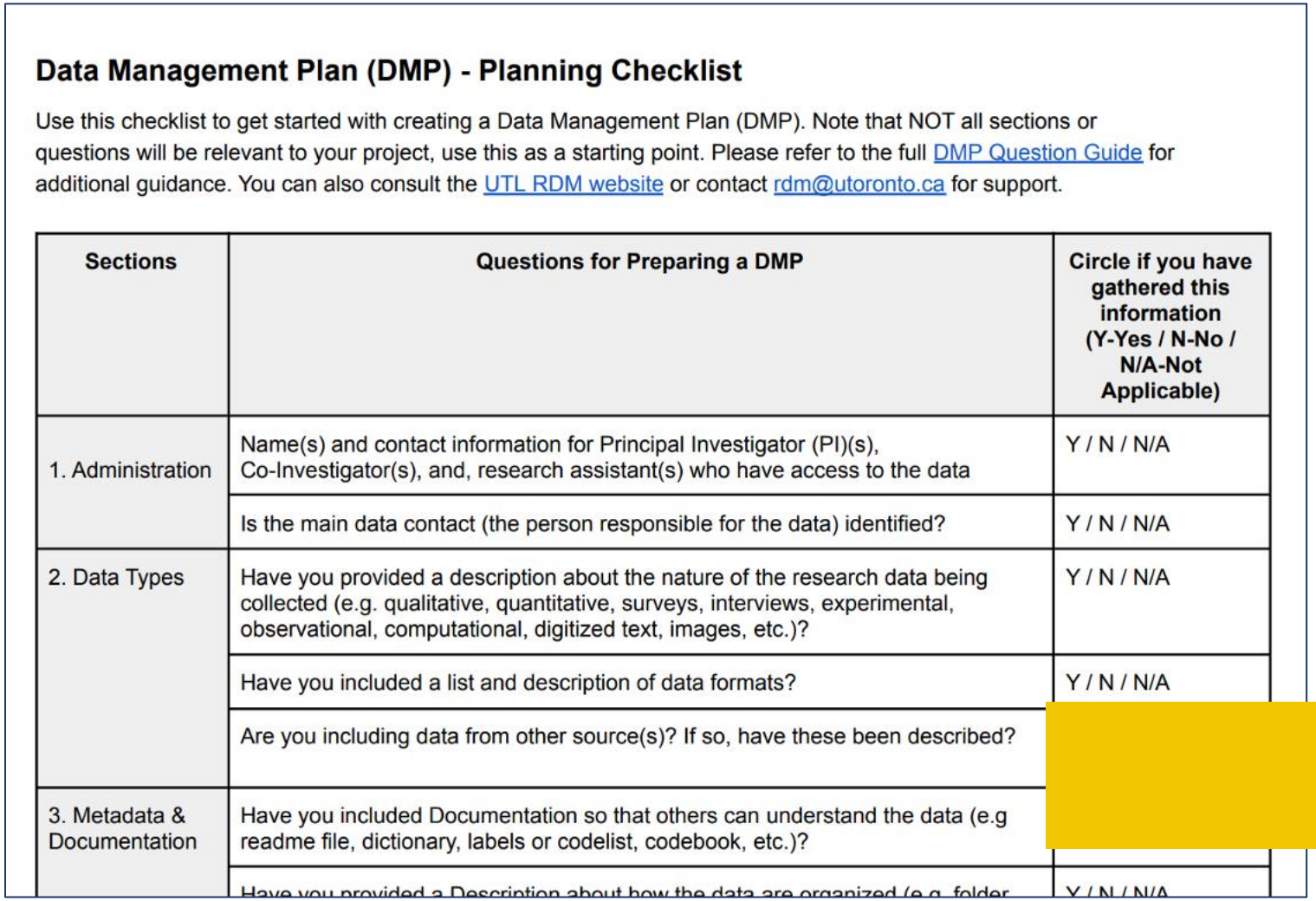
DRI Portal



The DRI Portal and the UTL RDM website offer best practices and resources related to data management planning

https://library.utoronto.ca/researchdata
https://cris.utoronto.ca/dri_portal/home/

UTL DMP guidance and checklist documents



Data Management Plan (DMP) - Planning Checklist

Use this checklist to get started with creating a Data Management Plan (DMP). Note that NOT all sections or questions will be relevant to your project, use this as a starting point. Please refer to the full DMP Question Guide for additional guidance. You can also consult the UTL RDM website or contact rdm@utoronto.ca for support.

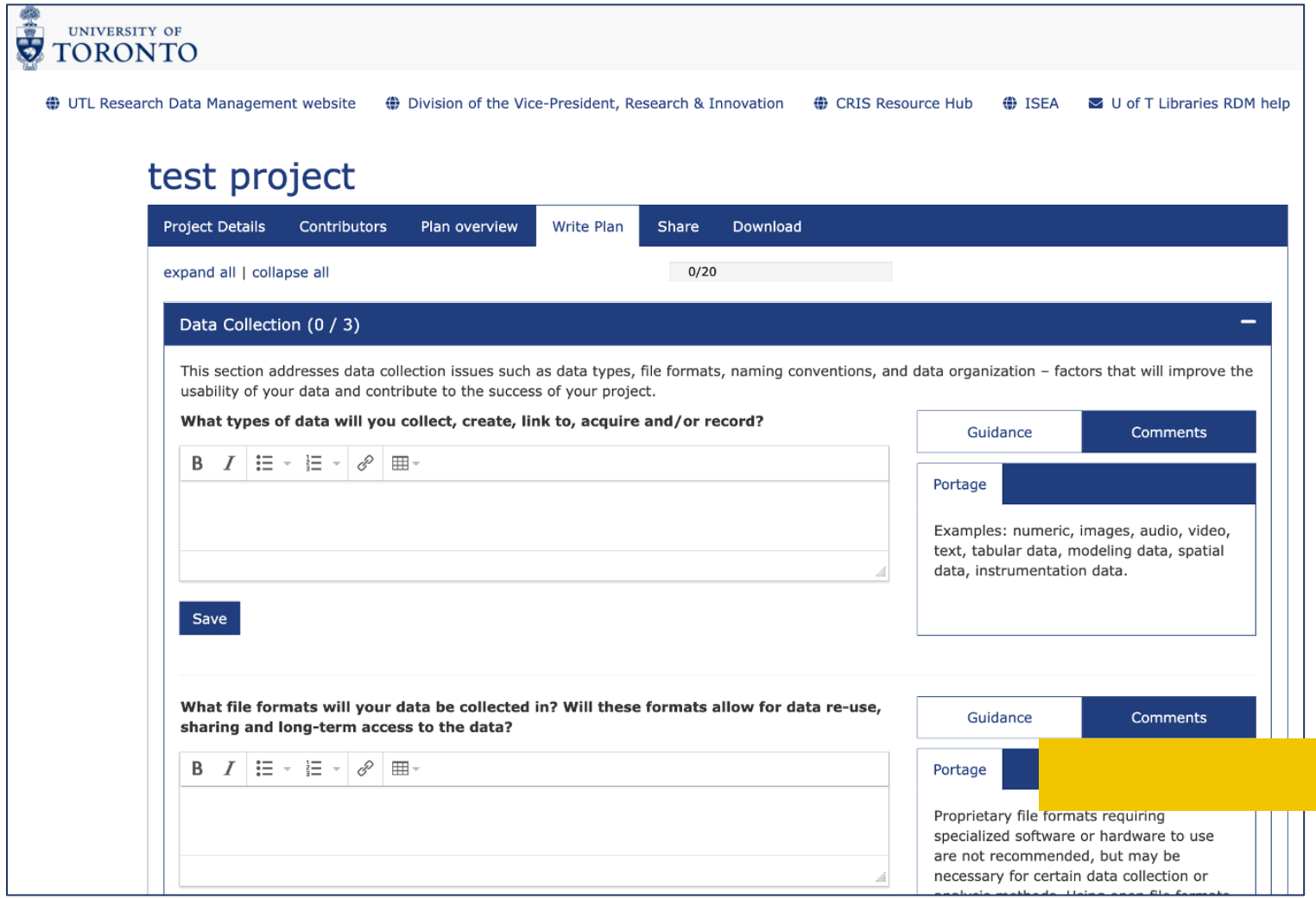
Sections	Questions for Preparing a DMP	Circle if you have gathered this information (Y-Yes / N-No / N/A-Not Applicable)
1. Administration	Name(s) and contact information for Principal Investigator (PI)(s), Co-Investigator(s), and, research assistant(s) who have access to the data	Y / N / N/A
	Is the main data contact (the person responsible for the data) identified?	Y / N / N/A
2. Data Types	Have you provided a description about the nature of the research data being collected (e.g. qualitative, quantitative, surveys, interviews, experimental, observational, computational, digitized text, images, etc.)?	Y / N / N/A
	Have you included a list and description of data formats?	Y / N / N/A
	Are you including data from other source(s)? If so, have these been described?	
3. Metadata & Documentation	Have you included Documentation so that others can understand the data (e.g readme file, dictionary, labels or codelist, codebook, etc.)?	
	Have you provided a Description about how the data are organized (e.g. folder	Y / N / N/A

UTL maintains resources that assist in preparing researchers to address typical DMP topics and questions

UTL DMP - Question Guide

UTL DMP - Planning Checklist

DMP Assistant

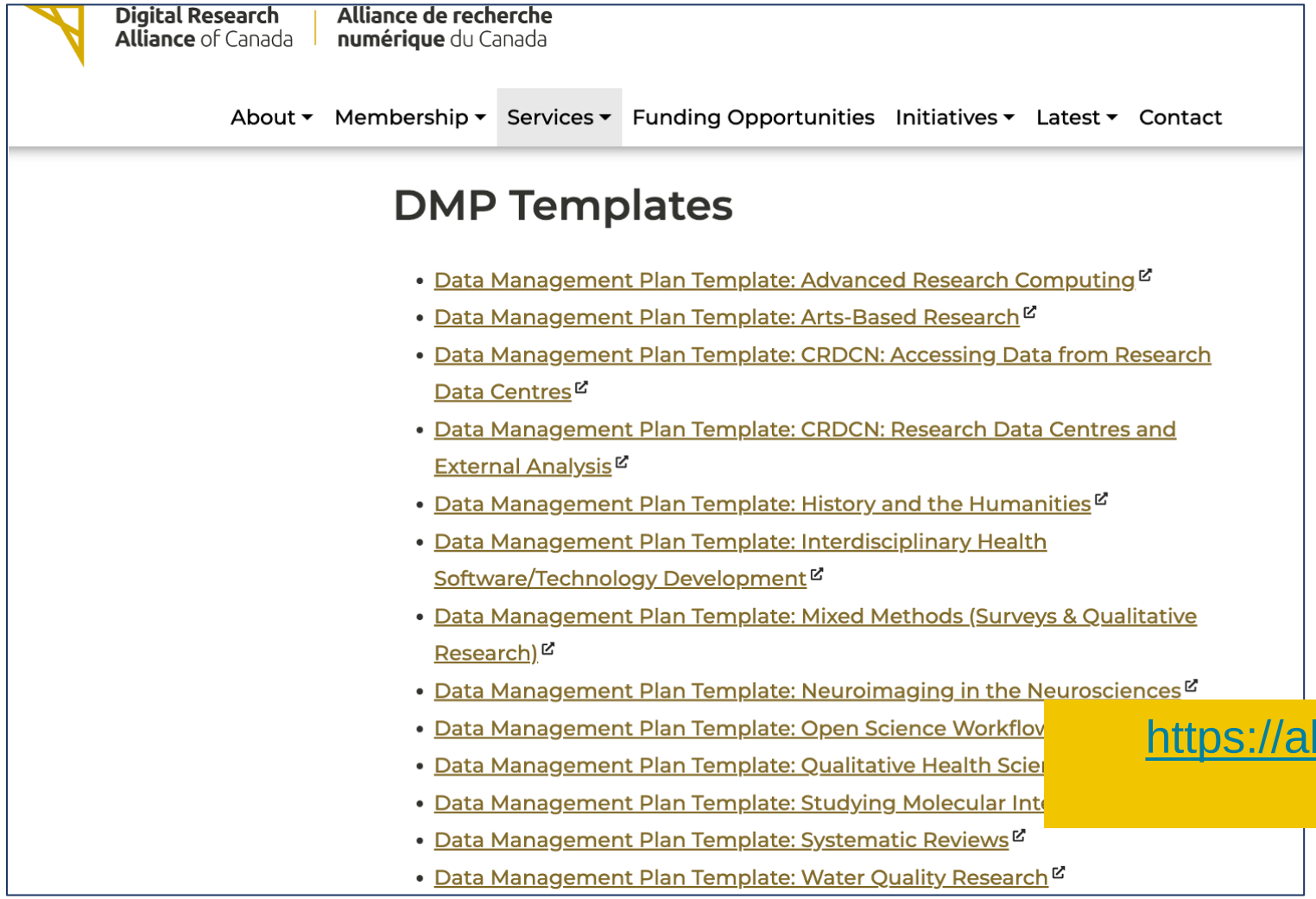


DMP Assistant is a national, bilingual tool available to assist researchers in preparing data management plans

- Includes templates, guidance and examples
- Freely available
- Supports cross-institutional research collaborations
- Role-based permissions (co-owner, editor, or read only)
- Exports to multiple formats

https://dmp-pgd.ca/

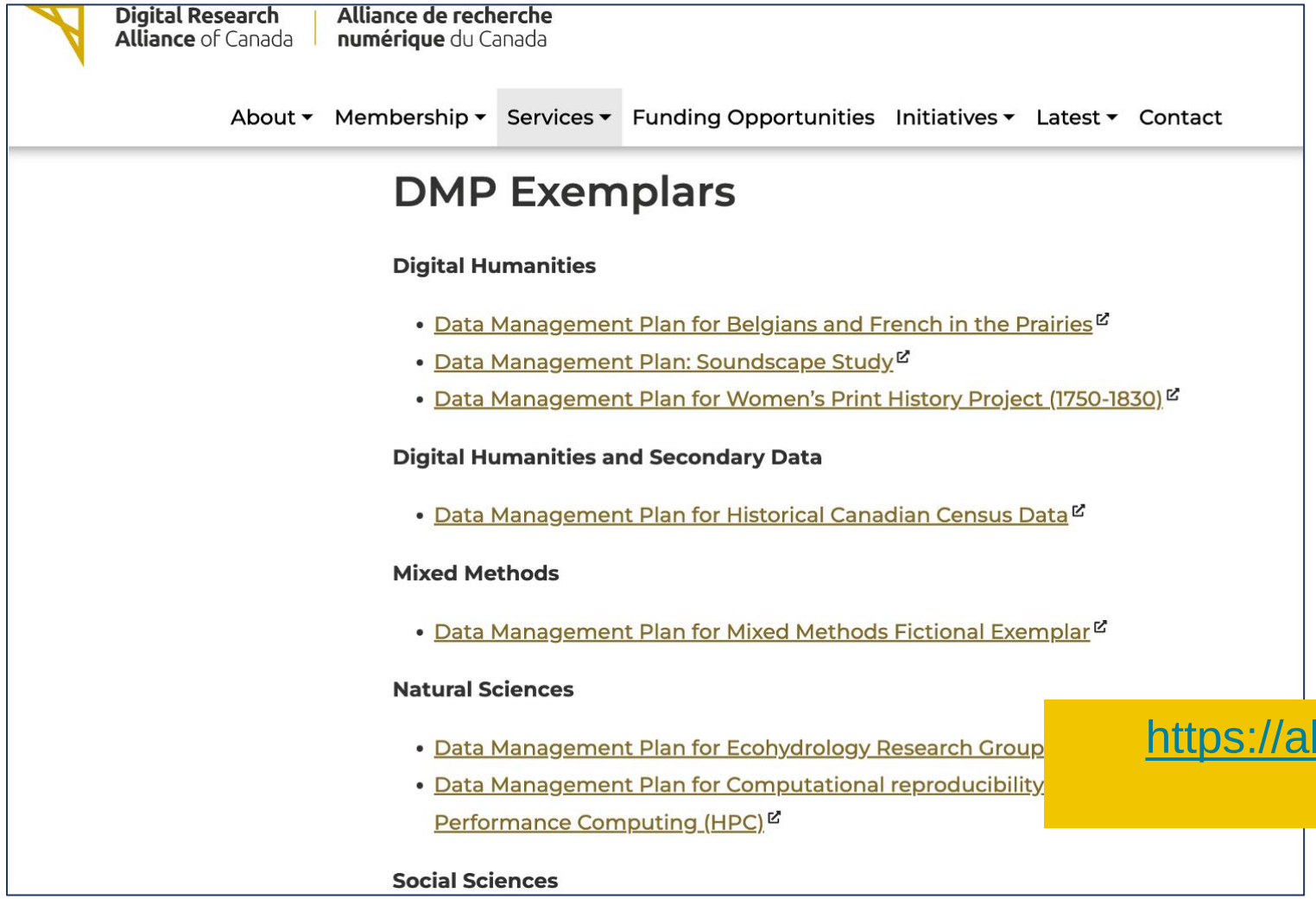
The Alliance – DMP Templates



The Digital Research Alliance of Canada offers DMP templates by discipline and research method

https://alliancecan.ca/en/services/research-data-management

The Alliance – DMP Exemplars



The Digital Research Alliance of Canada offers examples of DMPs created for different types of projects

https://alliancecan.ca/en/services/research-data-management

Help with DMPs

Services to support DMP creation and RDM practices are available through the Map and Data Library

- General questions
- Consultations
- Workshops
- Instruction sessions

https://library.utoronto.ca/researchdata
email: rdm@utoronto.ca or cris@utoronto.ca

Key Takeaways

Key takeaways

- DMPs can be a useful tool to help design an effective research project
- There are resources to help you create a DMP
- Plan for, and practice, good research data management - it will help your research team to:
  - Stay organized,
  - Improve the quality of your research,
  - Maintain access to your data over time,
  - Share your data with others, and
  - Fulfill funding, partner, and publisher requirements

Questions?

cris@utoronto.ca or rdm@utoronto.ca

UNIVERSITY OF TORONTO

DEFY GRAVITY

Acknowledgements

Some content modified from:

- Dearborn, D., Forbes, S., Umetsubo, Y., Henshilwood, A., Maistrovskaya, M. (2021). Research Data Management Level 2: Data Management Plans. University of Toronto Libraries.
- Grenier, M. (2021). Introduction to Data Management Plans. University of Victoria Libraries Digital Scholarship Commons.