

Tri-Agency Research Data Management Policy & Implications for Ethics Boards

Jay Brodeur, McMaster University Library

MREB | 18-November, 2019





Tri-Agency RDM Policy: Background & Summary



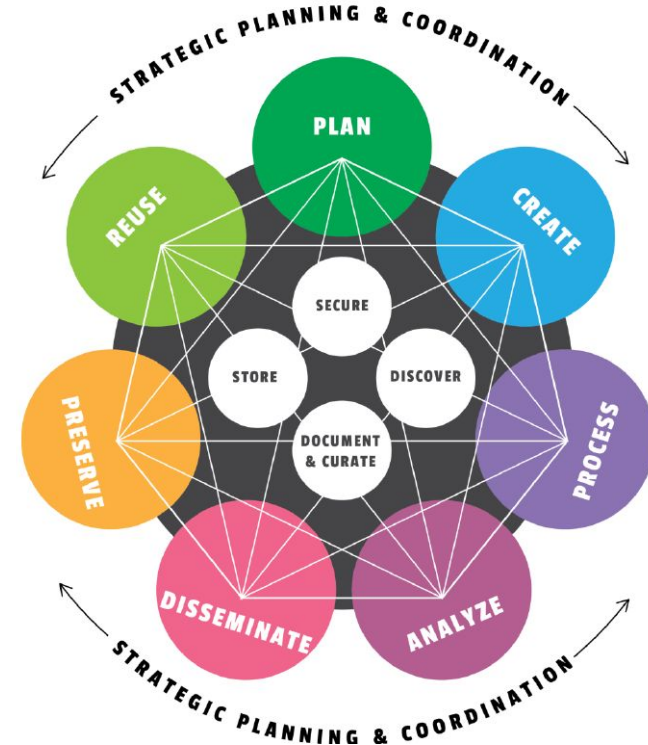
Research Data Management

Research data are:

- used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity,
- used as evidence in the research process and/or are commonly accepted in the research community as necessary to validate research findings and results

Research data management (RDM):

- includes the collection, storage, preservation and, where appropriate, access to data produced from a given investigation.
- should be practiced over the entire lifecycle of the data



Tri-Agency Statement of Principles on Digital Data Management¹ (2015 / 2016)

The screenshot shows the Science.gc.ca website. The header includes the Government of Canada logo and navigation links for Services, Departments, and Français. The main navigation bar features links for Partnerships and Projects, Science Links, Science News, and Educational Resources. The breadcrumb trail reads: Home > Research Funding Collaboration > Policies and Guidelines > Data Management > Tri-Agency Statement of Principles on Digital Data Management. The page title is "Tri-Agency Statement of Principles on Digital Data Management". The left sidebar lists various programs and policies, including "Meet Canada's scientists and professionals", "Canadian High Arctic Research Station", "Northern Contaminants Program", "Canada's Extended Continental Shelf Program", "Health Canada: Centre for Vaccine Evaluation", "Science and Technology Data Research Funding Collaboration", "Co-operative Funding Programs", "Policies and Guidelines", "Open Access", "Selecting the Appropriate Federal Granting Agency", "Research Ethics", "Responsible Conduct of Research", and "Public Communications". The main content area starts with "1. Preamble" and contains text about 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). It states that these agencies are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada. It also mentions that the agencies are strong advocates for making the results of the research they fund as accessible as possible.

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

Tri-Agency Statement of Principles on Digital Data Management: *Expectations*

- Data management planning
- Constraints and obligations
- Adherence to standards
- Collection and storage
- Metadata
- Preservation, retention and sharing
- Timeliness
- Acknowledgement and citation
- Efficient and cost-effective

Tri-Agency Statement of Principles on Digital Data Management: *Responsibilities*

Researchers

- incorporating best practices
- developing DMPs
- adhering to policies and standards

Research Institutions

- supporting best practices
- providing access to resources
- creating guidance and policies

Research Communities

- developing & promoting standards
- fostering excellence
- selecting repositories

Research Funders

- developing policies & guidance
- promoting data management
- providing peer reviewers



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

Draft Tri-Agency Research Data Management Policy² (2018)

 Government of Canada / Gouvernement du Canada



Home → Collaboration between Federal Research Funding Organizations → Policies and Guidelines
→ [Research Data Management](#)

DRAFT
**Tri-Agency Research Data Management Policy
For Consultation**

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 domain 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 are data that are used as primary sources to support technical or scientific enquiry, research, scholarship, or artistic activity, and that are used as evidence in the research process or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data or repurposed data. ^[1] Research data enable researchers to ask new questions, pursue novel research programs, test alternative hypotheses,

[2] http://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html [shortened: bit.ly/TA-RDM-Policy]



Draft Tri-Agency RDM Policy²

Released in June 2018; consultation period from June-Sep, 2018

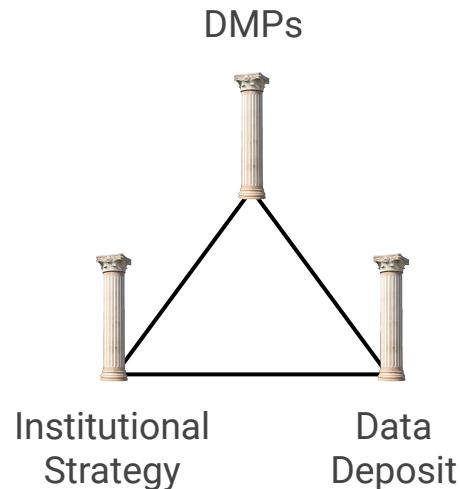
Will apply to Tri-Agency grant recipients and institutions administering tri-agency funds.

Three Pillars:

1. Institutional Strategy
2. Data Management Plans
3. Data Deposit

Planned launch in winter 2019.

Phased and incremental implementation



[2] http://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html [shortened: bit.ly/TA-RDM-Policy]



Institutional Strategy

“Each institution administering tri-agency funds is required to create an institutional research data management strategy. The strategy will outline how the institution will provide its researchers with an environment that enables and supports world-class research data management practices...”

- Fostering and promoting excellence in data management
- Supporting data management practices that are consistent with ethical, legal and commercial obligations [notably, the TCPS2]
- Providing guidance inline with **Tri-Agency Statement of Principles on Digital Data Management**¹

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



Institutional Strategy

“Each institution administering tri-agency funds is required to create an institutional research data management strategy. The strategy will outline how the institution will provide its researchers with an environment that enables and supports world-class research data management practices...”

- Developing their own data management policies and standards
- Ensuring that their researchers have data management plans in place
- Providing, or supporting access to, recognized repository services
- Consulting with existing resources to support the development of their strategies, such as the **Portage Network’s Institutional Strategy Template**³.

[3] <https://portagenetwork.ca/network-of-experts/institutional-rdm-strategy/>



Data Management Plans

“...The agencies encourage grant applicants to complete data management plans (DMPs) as an essential step in research project design. For specific funding opportunities, the agencies may require DMPs to be submitted to the appropriate agency at time of application; in these cases, the DMPs may be considered in the adjudication process.”



Data Management Plans

DMPs **are living documents** that should outline:

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

Grant applicants are encouraged to use standardized tools to develop their DMPs, such as the **Portage Network's DMP Assistant**⁴

[4] <https://portagenetwork.ca/network-of-experts/institutional-rdm-strategy/>



Data Deposit

*“Grant recipients are required to deposit into a **recognized digital repository** all digital research **data, metadata and code** that **directly support the research conclusions** in journal publications, pre-prints, and other research outputs that arise from agency-supported research...”*

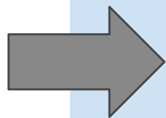
- The repository will ensure safe data storage, preservation, and curation
- The agencies encourage researchers to provide access to the data where ethical, legal, and commercial requirements [e.g. TCPS 2] allow, and in accordance with the standards of their disciplines.
- Whenever possible, these data, metadata and code should be linked to the publication with a persistent digital identifier.



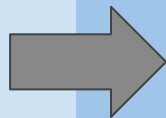
Points of Discussion



**Data in
Planning**

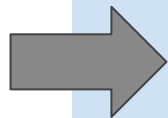


**Data in
Motion**

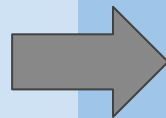


**Data
at Rest**

**Data in
Planning**



**Data in
Motion**



**Data
at Rest**



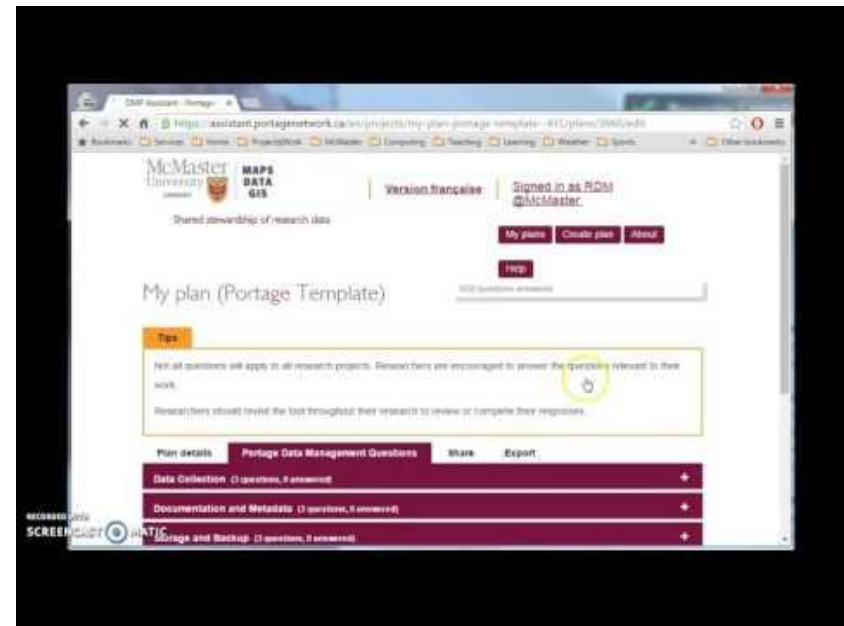
**Data in
Reuse**

Are DMPs asking the “right” questions?

Can researchers access resources to complete DMPs appropriately?

Does the institution have the capacity to evaluate DMPs?

How can the institution streamline DMP and REB processes?



Portage DMP Assistant
assistant.portagenetwork.ca

**Data in
Planning**



**Data in
Motion**



**Data
at Rest**

Are we providing researchers with appropriate:

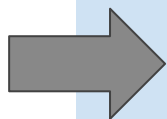
- Triage and assessment guidance
- Methodological guidance
- Options for file storage, backup, sharing?
 - Deeper consideration of cloud options like OneDrive and sync.com

How to enable sharing across institutional collaborators that satisfies all REBs?

The screenshot shows the 'Data Storage Finder' tool. At the top, it says 'Evaluate options for data storage at Cornell'. Below this are three green icons with text: 'All services presented on this finder tool are vetted and supported by Cornell University.', 'To explore data options available to Weill Cornell Medicine Cornellians please visit the [WCMC storage wizard](#).', and 'We welcome [feedback](#) on this tool.' The main section is titled 'Describe your data'. It contains instructions: 'Answer these questions to help identify data storage services that are suitable for your needs. Checking these boxes will change the list of available services. If you are uncertain how to answer, leave the question blank to maximize your resulting options.' There is a 'Clear Answers' button. Below this are two questions: '1. What is the classification of your data?' with radio button options for 'Public', 'Sensitive / Moderate Risk', 'Confidential or Restricted / High Risk', and 'HIPAA-Regulated'; and '2. Do you need backups, snapshots or replication of your data?'. To the right of these questions is a grid of storage service options, each with a radio button. The services listed are: Amazon Web Services Elastic Block Store, Amazon Web Services Elastic File System, Amazon Web Services Glacier, Amazon Web Services S3, BioHPC Cloud, CAC Archival Storage, CAC Red Cloud Storage, CISER Research Servers, CUGIR, CUL eCommons, Cornell Box, and Cornell Restricted Access Data Center. At the top right of the grid are 'Select All' and 'Clear Selections' buttons.

Cornell Data Storage Finder
finder.research.cornell.edu/storage

Data in
Planning



Data in
Motion



Data
at Rest

Moving beyond 'default = destroy' ...

... while assessing re-identification risk

Understanding and accommodating
'disciplinary standards and practices'

Demystifying repository vs archival storage

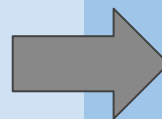
Portage Responsible RDM Practices
for Sensitive Data Working Group

<https://portagenetwork.ca/network-of-experts/responsible-rdm-practices-for-sensitive-data-working-group/>

**Data in
Planning**



**Data in
Motion**



**Data
at Rest**



More Information

McMaster Library's RDM page

library.mcmaster.ca/services/rdm

Contact: rdmgmt@mcmaster.ca

CARL Portage Network

portagenetwork.ca



Library

RDM
@McMaster