

## Portage Comments on Draft Tri-Agency Research Data Management Policy

Original Tri-Agency [DRAFT POLICY](#) & [FAQ](#)

### Background

The Portage Network, an initiative of the Canadian Association of Research Libraries (CARL), is dedicated to the advancement of research data management (RDM) in Canada. Portage, through its nation-wide network of over 115 experts from over 40 organizations, works proactively with community partners to develop national RDM services and infrastructure in support of all researchers in Canada.

Portage had the privilege of working closely with Tri-Agency representatives during the pre-consultation phase of policy development, traveling the country to meet with 14 different stakeholder groups in 4 cities in Jan-Feb 2018. These meetings provided an ideal opportunity both to share Tri-Agency policy requirements with researchers and to describe Portage services and infrastructure designed to help researchers fulfill each requirement. Additionally, these meetings gave Portage, alongside our Tri-Agency colleagues, the chance to hear feedback from stakeholder groups and to respond together to questions that arose.

Portage also took the opportunity to convene three Town Hall meetings in July 2018, engaging over 32 participants. Feedback from these Town Halls and other less formal discussions was considered when crafting the Portage response provided below. A summary of comments from the community is provided in **Appendix A**.

### Introduction

Portage appreciates the thought and hard work put into formulating the Tri-Agency draft RDM policy. Portage fully supports the Tri-Agency in its RDM Policy directions and welcomes the opportunity to provide feedback during this public consultation period. Comments are organized below, following the sections of the draft policy.

### Specific Comments Related to the Policy:

#### 1. Preamble

Suggested changes shown in **underscored red**, below:

“Data management plans assist researchers in determining the **logistics**, costs, benefits and challenges **associated with** managing data and should be consulted and updated throughout the research project.”

“The Government of Canada’s [Action Plan on Open Government](#) aims to maximize access to the results of **federally-funded** research, to encourage greater collaboration and engagement with the scientific community, private sector and public.”

COMMENTS: The preamble provides a strong introduction to and endorsement of the

concept of RDM and key relationships with other government policies and policy directions. The preamble uses phrase like:

- 'support research excellence'
- makes 'research results as accessible as possible'
- 'research data has become critical...'
- 'research data management is necessary'
- DMPS are 'an international best practice' and 'integral to research excellence'
- 'need for policies to enable excellence in data management'
- 'maximize access to the results of federally-funded research'
- 'adopt policies to support effective data stewardship'
- 'promote excellence in data management practices'

The community felt that the draft policy did not entirely live up to these expressions of support for RDM and therefore that these ideals should be reflected in a firmer and more direct set of policy requirements.

## 2. Policy Objective

No comments.

## 3. Policy Statement

ORIGINAL TEXT: 'This policy applies to grant recipients and to institutions administering tri-agency funds. It does not apply to scholarship, fellowship or Chair holders'

COMMENTS: There is a logical inconsistency here that will not be lost on grant seekers obliged to follow the new policy. The fact the draft policy excludes 'scholarship, fellowship or Chair holders' needs to be explained. If the intent is that these recipients of Tri-Agency funds will be eventually folded into the policy, this should be stated.

ORIGINAL TEXT: 'This policy currently only applies to grant recipients and to institutions administering tri-agency funds. It does not yet apply to scholarship, fellowship or Chair holders.'

### 3.1 Institutional Strategy

ORIGINAL TEXT: 'committing to develop their own data management policies and standards for data management plans'

COMMENTS: While we acknowledge the need for institutions to address local disciplinary and administrative needs, allowing each institution to '*develop their own data management policies and standards*' is too open-ended. Best practices surrounding DMPs are well-established -- see the Portage [DMP Assistant](#) for an

example of how these best practices are expressed in an online tool. These best practices should set the standard for DMPs and inform any eventual institutional policies that are developed. Leaving this up to institutions could lead to watered-down DMP requirements, from both a content and compliance perspective.

SUGGESTED TEXT:

'committing to the adoption of established best practices when developing institutional standards and policies for data management plans.'

ORIGINAL TEXT: 'providing, or supporting access to, recognized repository services or other platforms that securely preserve, curate and provide continued access to research data'

COMMENTS: This should read 'providing continued appropriate access' to acknowledge the fact that not all data can be made openly accessible.

SUGGESTED TEXT:

'...provide continued appropriate access to research data'

## 3.2 Data Management Plans

### ***Submission requirements***

ORIGINAL TEXT: All grant proposals submitted to the agencies should include methodologies that reflect best practices in research data management. 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.

COMMENTS: The DMP requirement needs to be clear and direct. The current draft's use of '*should include*', '*encouraged*', '*may require*' and '*may be considered*' falls short in this regard, particularly when DMPs are described in the same paragraph as being '*an essential step in research project design*'. Further, the policy preamble states that DMPs are '*integral to research excellence*'. If they are '*essential*' and '*integral*', the policy's requirements for DMP submission, evaluation, and updating should be strengthened accordingly.

### ***Adjudication***

COMMENTS: The Tri-Agencies should clarify their vision for DMP adjudication in terms of when DMPs should be submitted and updated (*DMPs are 'living documents'*), criteria used to evaluate DMPs at various stages in the research lifecycle, and who will conduct these reviews. The EU's Horizon 2020 policy states: "*Principal Investigator must develop a data management plan in the first 6 months of the project and keep it up-to-date throughout their project.*".

SUGGESTED TEXT:

"All grant proposals submitted to the agencies must include data management plans (DMPs) as an essential step in research project design. DMPs will be considered in all funding applications and, for specified funding opportunities, will be evaluated as part of the adjudication process. Principal Investigators on approved projects are responsible for ensuring DMPs are kept up-to-date throughout the project."

### 3.3 Data Deposit

#### ***Definition of what data needs to be deposited***

ORIGINAL TEXT: "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."

COMMENTS: This text could be used to drastically limit what data gets deposited. What is described here is known as 'replication data' -- the strict minimum of data needed to replicate a table, chart, or regression (for example) in a published paper.

As written, researchers could, for instance deposit data for the 5 variables used in a regression, or a chart. All other variables, which provide context and value to the overall research data set, could be excluded under the current language.

Portage feels there is value in broadening the definition to include all final data, documentation, and code associated with a research project, not just the subset used in a published article. Portage believes this was the Tri-Agency's intent -- but this needs to be made clearer in the policy.

Further, the [ESRC](#) (Economic and Social Research Council) in the UK requires grant holders to deposit data suitable for '*preservation and future reuse*'. In the context of deposit into a repository, *suitable for preservation* denotes the best practice of using non-proprietary, platform-agnostic formats that support eventual preservation/archiving/forward migration of the data, should the curatorial decision to do so be made. *Suitable for future reuse* speaks to the importance of ensuring data, metadata, and code deposited in a repository are complete and of sufficient quality to be independently understandable by future researchers.

SUGGESTED TEXT:

"Grant recipients are required to deposit all digital research data, metadata and code associated with research outputs arising from agency-supported research (i.e. published journal publications, pre-prints, and other research outputs) into a recognized digital repository in a form that supports both preservation and future reuse."

***Repositories: ‘preservation’ and ‘curation’***

ORIGINAL TEXT: “The repository will ensure safe storage, preservation, and curation of the data.

COMMENTS: *Repository storage* is intended primarily for discovery and access, and under RDM best practices would certainly include curation and safe storage of data, documentation, and code. Curators may decide that *some* data housed in a repository should be specially treated for long-term preservation in *Archival storage*. Preservation in this context denotes the application of special processes and practices that ensure the long-term storage and maintenance of data, documentation, and code, and is not the purview of data repositories.

There is some question about what is meant by ‘curation’ and how this requirement will apply to repositories. For example, there are repositories that offer researcher ‘self-deposit’ -- with no curatorial role provided or assumed. Curation is a complex and involved undertaking for any repository; there is need for clarity on who will take on this role and how it will be funded.

SUGGESTED TEXT:

“The repository will be responsible for the curation and safe storage of deposited data, as well as for facilitating discovery and appropriate access.”

***Data access, timing of deposit, embargos, and linkage***

ORIGINAL TEXT: “The agencies encourage researchers to provide access to the data where ethical, legal, and commercial requirements 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.”

COMMENTS:

The current draft policy makes no mention of when researchers are expected to deposit their data. The UK [Concordat on Open Research Data](#) states that “*Researchers will, wherever possible, make their research data open and usable within a short and well-defined period, which may vary by subject and disciplinary area and reflect the resources available to them to do so. Data supporting publications should be accessible by the publication date and should be in a citable form. Where it is not possible to make data open for legitimate reasons, there should be no negative consequences for those researchers concerned.*” The [ESRC](#) (Economic and Social Research Council) in the UK requires grant holders to deposit data ‘*within three months of the end of the grant*’.

SUGGESTED TEXT:

“Researchers will, whenever possible, make their research data open and usable within a short and well-defined period, based on disciplinary norms. In some instances, delaying access to data (i.e. a clearly defined data embargo period) may be justified. For data that cannot be made fully open for ethical, legal, commercial,

or other legitimate reasons, metadata and code must be made open and data access agreements put in place to define and facilitate appropriate controlled access to the associated data. Data, metadata and code housed in a repository should be linked to research outputs such as publications with a persistent digital identifier.”

#### **4. Implementation Date**

There is considerable concern and interest in the community about when and how the policy will be implemented. Planning at the institutional, regional, and national level will be influenced by the roll-out schedule for this policy. Every effort should be made to make this public as soon as possible.

#### **5. Compliance with Policy**

ORIGINAL TEXT: “By accepting agency funds, institutions and researchers accept the terms and conditions as set out in the agencies’ policies, agreements and guidelines. In the event of an alleged breach of agency policy, agreement or guideline, the agency may take steps outlined in accordance with the [Tri-Agency Framework: Responsible Conduct of Research](#) to deal with the allegation.”

COMMENTS: More information is required here about how compliance will be monitored and enforced. Who will monitor compliance and how? What are the consequences of non-compliance for a researcher or institution? Who will pay for all of this? This information should be stated explicitly.

#### **6. Policy Review**

No comments.

#### **7. Additional Information**

No comments.

## Appendix A. Overarching Comments/Questions from the Community

-- based on Town Halls and other informal discussions:

- Give institutions and organizations trying to build infrastructure and services around RDM a stronger mandate; makes it easier for us to sell RDM at our institutions and to obtain funding to support these initiatives.
- It will be easier for institutions to create institutional RDM strategies and eventual RDM policies if the Tri-Agency policy is more stringent in its policy requirements.
- Some concern about institutions across Canada having different policies/standards on RDM -- inconsistencies could cause problem with cross-institution research projects and/or researchers moving from one institution to another.
- Infrastructure to support these initiatives requires investment in DRI (*and specifically RDM*)
- Funders need to take an active role in administering and ensuring compliance with this policy.
- What funding and support will the Tri-Agencies provide to help researchers and institutions comply with the policy?
- We need a clearer statement on how compliance will be monitored and enforced (e.g. include DOI to deposited data at the end of a grant?). Who will do this and who will pay?
- The Tri-Agencies should explicitly state (in FAQ?) the importance of metadata and code in supporting data sharing and reuse.
- Emphasize the importance of exposing metadata and code even if restrictions apply to accessing the actual data.
- DMPs should flow through the Tri-Agencies, and be a real part of the granting/research life cycle -- including a requirement for updates to DMPs during the grant period.
- There needs to be more emphasis on DMPs being 'living documents' needing updating throughout the research life cycle.
- There needs to be more information/clarity on the issues of students, Canada Research Chairs, etc. being excluded from the policy as currently written.
- We need more information on timing and pace of the policy roll-out.
- There was some interest in seeing the RDM policy align more closely with the government's Open Data policy. While not an 'open data' policy, the policy should increase emphasis on the value of data sharing.
- Some concern about what is meant by 'recognized digital repository' -- work needs to be done on 'trusted' status (a 'seal of approval'?) for repositories.
- Tri-Agency needs to be clearer on when data needs to be deposited and mechanisms to address researcher concerns (e.g. data access agreements, embargoes).