Response to the Draft Tri-Agency Research Data Management Policy For Consultation

AUGUST 2018

**Office of the Vice-President
(Research and International)**

UNIVERSITY OF MANITOBA

CONTENTS

BACKGROUND

The following document is a single, consolidated response by the University of Manitoba (UM) to the [Draft Tri-Agency Research Data Management Policy](#),[1] which aims to support Canadian research excellence by fostering sound digital data management and data stewardship practices. The draft policy, developed by 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), applies to grant recipients and to institutions administering Tri-Agency funds. It includes suggested requirements related to 3 primary areas: Institutional Data Management Strategies, Researcher Data Management Plans, and Data Deposit. [The Tri-Agencies have asked for feedback](#) on the Draft Policy to inform its final design, which the agencies plan to launch in 2019.[2]

The UM is appreciative of the Agencies' consultative approach on this important initiative. The UM strongly supports the Agencies' recognition of the importance of research data management (RDM) for research excellence at all stages of the research lifecycle, as demonstrated by this draft Policy, as well as previous statements and policies.

The University of Manitoba hosted two Town Halls in August to gather feedback from the UM research community. Researchers, administrators and other stakeholders were also encouraged to submit their feedback [via a UM online form](#).[3] The discussions focused on two key questions:

1. How will the data be managed?
2. How will this policy be implemented?

DISCUSSION SUMMARY

Question 1: Data Management Plan
What is being proposed requires improved clarity and transparency. The Policy's requirements for data management plan (DMP) submission, evaluation, and updating should be strengthened. The use of 'should include', 'encouraged', 'may require' and 'may be considered' need to be reconsidered if DMPs are truly 'essential' and 'integral', as stated in the Preamble. Similar to the Tri-Council Policy Statement 2 (TCPS 2, 2014),[4] there could be examples included to help illustrate specific concerns and provide interpretation support.

Consent
Careful consideration of encryption, de-identification, and re-identification of research data will be necessary to mitigate risks. Where there are potential questions around re-identification, there requires recognized direction for access, disclosure and use. In many research projects, consent is a significant and important component – how consent and data management interact is often problematic (examples of withdrawal of data, length of storage and where, governance of use, etc.). If researchers are required to anonymize the identity of their participants, some of the data may make re-identification possible and data may need to be

removed. Doing so may render the data ineffective. The TCPS 2 has some general guidelines about potentially identifiable data at rest and data in use with regard to encryption and storage/use on networks that are problematic in interpretation.[4]

Concerns were also raised regarding the privacy of personal health data and intellectual property, which may have stipulations regarding patent timelines. An 'opt-out' option could be provided for researchers, concerning certain aspects of the data management plan that could otherwise violate their terms and agreements.

Consented data collection can also have implications with regard to digital repositories, especially data collection that pre-dates more recent policies. Therefore, data linkage potential and limitations should be included in the data management plans.

Data Deposit

In order to support future reuse (which requires that data, metadata, and code deposited in a repository are complete and of sufficient quality to be independently understandable by future researchers), we would like to see a broadening of the definition of deposited data to include all final data, documentation, and code associated with a research project, not just the subset used in a published article. The format of code developed with protected intellectual property from industrial partners will also need to be addressed. Field data adds another dimension, for which Agriculture and Agri-Food Canada's Aquarius could be a valuable case study.

Additionally, the current draft Policy makes no mention of when researchers are expected to deposit their data. We recommend language that specifies that data be deposited within a short and well-defined period, based on disciplinary norms. Finally, the repository responsibilities listed by the Policy require clarification. While some repositories provide deposit and curation services, others rely on self-deposit. Parameters should be stipulated or at least projected, including the window of conversion, its size, the level of "raw" data to be retained (data granularity), and reasonable terms of expectations of data reuse. The inclusion of source and access for researchers and review bodies should also be recognized. The guidelines must be clear and concise, as any additional requirement regarding research may be unnecessarily viewed as a bureaucratic burden to academic and administrative workloads.

Additional questions that were raised included:

- Data sets from Proprietary software (of which there can be no 'open' conversion)
- Does the data repository need to be geographically located in Canada? Some of these large storage services can vary significantly in cost. A Canadian Federally-funded product called the Federated Research Data Repository (FRDR) seeks to provide data deposit for large datasets, but there is a concern that a new Government could withdraw funding for these types of repository services
- If data deposit is centralized, different disciplines have differing costs: large data sets are cost prohibitive and onerous for a data manager to navigate (constantly negotiating different standards, sizes).

- What role do large corporations such as Amazon and Google play in this Canadian research environment given that they are the largest services providers of digital storage?

Data Reuse

Data retention and archiving is an important part of research. What will be reasonable in terms of expectations of data reuse? Some useful analysis may come many years after the field data collection has been completed and the project has ended. In addition, open access to research data could deprive the original researchers from getting more publications out of the data that they meticulously collected.

Another possible issue with re-use is that the way in which the data was collected may make it such that the data is not valuable to others. It was collected for a specific study; it may not be possible to re-use that data for a future study. In such cases, does the data still need to be deposited and made available for others. The Draft Policy could encourage researchers/institutions to work together and collaborate on projects, which may be a more effective way of ensuring re-use of data (rather than having data uploaded to a repository).

Data Access

Data sensitivity deposit options will need to be established. For instance, this should include a process to seek input for and honouring Indigenous-related data. OCAP Principles[5] will need to be honoured.

Journals requiring the use of these deposits should clearly identify the owner and the researchers. What are the expectations between reuse protection for potential future commercialization of intellectual property?

Data Preservation

Preservation of deposited datasets is currently not supported by many recognized repositories. Depending on various characteristics of the data, such as format, size and completeness, the burden of preservation may be very challenging and time consuming.

Question 2: Implementation

Planning at the institutional, regional, and national level will be affected by the roll-out schedule for this Policy. Institutions need to be provided with direction and support regarding the multiple facets of this policy, in order to maximize the efforts of the institutions, researchers and administrators. Not all institutions will have the same ability to provide resources to support the highest level of rigour in this area. Presumably the policy will recognize this and show how such variation can be accommodated and will be used to create a framework against which the institutional support in this area can be measured. Every effort should be made to make this public as soon as possible. Scenarios and cases, cross-referenced to the policy, would be helpful.

Furthermore, it would help to provide more context and transparency regarding who has requested this policy, and what they stand to gain or lose from its implementation.

Institutional Strategy

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 problematic. Best practices surrounding DMPs are well-established, and these should set the standard for developing institutional policies related to DMPs.

While developing an institutional strategy for DMPs will benefit institutions regardless of this Policy, doing so requires significant resources. We recommend that the Policy specifically address how the Tri-Agency plans to support institutions in completing this exercise, including referrals to leverage existing community infrastructure providers.

Additional questions that were raised included:

- Will there be an establishment of protocols arising from the Policy?
- How does open deposit receive attribution (for example, the ORCID initiative; products like Scopus and/or Dimensions) and will this be standardized?

Language

We encourage the Agencies to use more direct and clear language throughout the Policy, specifically by making requirements for data management plans (DMPs), data sharing, and other key elements more explicit, rather than suggestive. This would create a clear expectation, give institutions a stronger mandate to build RDM infrastructure and services, and make it easier to obtain the funding required to implement the proposed Policy.

Indigenous Knowledge

We are pleased to see Indigenous peoples referred to in the FAQs but would prefer to see reference to Indigenous people and communities' data and related guiding principles included in the main Policy statement. Relevant information regarding data management and research with Indigenous peoples and communities ought to be drawn from Chapter 9 of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS)[4] and feature prominently in the Policy, rather than segregated in the FAQs.

Incidences of reused or shared data without clear, prior and informed consent from Indigenous communities and individual Indigenous research participants are extremely problematic. We recommend that "Secondary Use of Information or Human Biological Materials Identifiable as Originating from Aboriginal Communities or Peoples" be referred to directly in the Policy to stress the importance of this information in developing data management plans and agreements in collaboration with Indigenous communities.

We also recommend that the Policy explicitly recognize that ownership of Indigenous traditional knowledge shared during a research project is retained by Indigenous people, families, and communities, and that ownership of Indigenous traditional knowledge cannot be transferred to researchers as research data (OCAP Principles).[5]

Timeline and Compliance

It is imperative that these changes be made in reasonable stages, as to allow for both the institution and its administrative staff to properly prepare and execute, prior to implementation

with the academic staff. A well-executed data management plan will naturally lead to higher levels of compliance. 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? This information should be stated explicitly.

CONCLUSIONS

This document details feedback from the UM on the Draft Tri-Agency Research Data Management Policy. Ultimately, there are some major concerns that accompany the potential mandating of both short and long-term data storage and access. If Canadian universities lack the capacity to manage the data themselves, then researchers are asked to trust their data with commercial entities, governments, or poorly funded small-scale services that may lack the capacity to provide the services they require. Where data storage and management is outsourced, research data is vulnerable to many risks that endanger the underlying mandate of university research to be independent and advance human knowledge. With government storage, data is vulnerable to accidental or deliberate deletion or alteration and other risks associated with changing political agendas and/or failures in competence or capacities. With commercial storage, data is equally vulnerable to these risks, in addition to risks associated with authorized or unauthorized storage outside of Canadian national boundaries. When data is stored and managed outside academe, it is vulnerable to many risks that pose threats to the independence of research as well as researchers' future ability to build on existing knowledge.

In closing, while the preservation of and access to data should be a priority for the Tri-Agency, the Tri-Agency should support Canadian universities to develop a harmonized set of data storage facilities and strategies that are managed at the level of the university. The mandate to preserve and provide access to data should be accompanied by an initiative to fund the development of expertise and capacity to preserve data and curate it long term within and across universities in Canada.

REFERENCES

[1]   Government of Canada (2018). *Draft Tri-Agency Research Data Management Policy for Consultation*. Retrieved from www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html. Retrieved on 2018-07-25.

[2]   Government of Canada (2018). *Research data management.* Retrieved from www.science.gc.ca/eic/site/063.nsf/eng/h_547652FB.html. Retrieved on 2018-07-25.

[3]   University of Manitoba (2018). *Tri-Agency research data management policy.* Retrieved from libguides.lib.umanitoba.ca/researchservices/tri-agency-feedback. Retrieved on 2018-07-25.

[4]   Government of Canada (2017). *TCPS2 (2014) – the latest edition of Tri-Council policy statement: Ethical conduct for research involving humans.* Retrieved from www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/Default/. Retrieved on 2018-07-25.

[5]   First Nations Information Governance Centre (2018). *The First Nations principles of OCAP*. Retrieved from fnigc.ca/ocapr.html. Retrieved on 2018-07-25.