



Draft Tri-Agency Research Data Management Policy University of Toronto Response September 2018

The University of Toronto (UofT) strongly supports both the implementation of research data management (RDM) practices and making Agency-funded research data available for reuse by others in a responsible way (i.e., considering factors such as commercial, ethical, and legal obligations, as well as costs and benefits). We appreciate that we have been asked to provide institutional feedback on the draft *Tri-Agency Policy on Research Data Management* (the *Policy*).

The *Policy's* requirement to have research data management plans (RDMPs) delineated within grant applications is timely. It is also appropriate that the *Policy* recognizes that the responsibilities for data management are evenly distributed between researchers, research communities, institutions and the Agencies, and that all stakeholders have a role to play in research data management. It is not clear why Research Chairs and other groups of scholars would be excluded from the *Policy* requirements.

We believe it will be important that the *Policy* be issued as soon as practicable to allow a longer lead time for full institutional and research compliance, so that stakeholders have time to put into place appropriate policies, support and guidelines that are based on the final *Policy* requirements. However, it will be essential that the *Policy* use a phased-in approach for having RDMPs (and data storage and access) and for requiring, under appropriate circumstances, the sharing of data with others. We propose that 3 years to fully implement the *Policy* would be appropriate.

A phased-in approach is very important as the *Policy* will require a change in current practice for many researchers. RDM practices are very field- and discipline-specific; some fields/disciplines are certainly already practicing RDM (e.g. in biomedical research and large data-intensive cohort studies), but some disciplines lag behind. RDM is currently not a scholarly norm and substantial training and support will be required. A key to the development of scholarly norms is discussion and debates within disciplines. Many scholarly associations are leading efforts to develop appropriate mechanisms for data management and curation. This is important work, but also takes time to develop the agreed-upon standards and common frameworks. For example, UofT researchers are part of the International Image Interoperability Framework (IIF) that is a collaborative effort to produce an interoperable technology and community framework for image delivery. The IIF formed in 2014 and has recently released drafts of interoperability specifications for review and discussion for the community.

The change in practice is not only field and discipline specific. Researchers collaborate across fields and disciplines and to work effectively they need to have similar practices for data management and data sharing. There needs to be some time for a scholarly norm to be developed for multi-disciplinary teams. Needing time to develop a change in practice is also true for researchers who enter into agreements with other academic institutions, private sector partners, and not-for-profit organizations both nationally and internationally. Institutions creating a strategy for research data management need time to consider the many existing RDM policies in place in other jurisdictions so that whatever standards are chosen, they are congruent with the existing RDM policies to ensure that data can be exchanged and used internationally.

In terms of data sharing requirements, this can only realistically be done once there is a change in current practice. There are many issues to be resolved such as the jurisdictional, legal, and ethical aspects of RDM before implementation of RDMPs for particular disciplines, especially around the ownership of research data, intellectual property and publication issues, and private and/or sensitive data. As an example, researchers have successfully demonstrated that small amounts of key data can be used to identify individuals from supposedly anonymous records. RDMPs will need to consider the requirements for anonymized data that is made available to ensure that the information cannot easily be rendered re-identifiable.

In terms of implementing the *Policy*, it is difficult for us to assess the implications to our institution and researchers without further clarity on some key *Policy* requirements:

Financial resources

For researchers, clarity is needed as to whether researchers can include appropriate resources to support data management for the duration of the grant as part of their budget. For example, effective data management and sharing requires support for time-consuming but necessary activities such as data cleaning, normalization and harmonization, etc. that grant peer review panels often do not support. For the *Policy* to be implemented, Tri-Agency-funded grants would need to include support for such management activities. For institutions, there are substantial institutional costs associated with this *Policy* including costs for the development of protocols, implementation of training and education, hiring of highly qualified support staff and the creation of infrastructure for the maintenance and storage of research data in their myriad of forms, whether institutionally or in disciplinary repositories. Additional funds should be allocated to allow for the uptake and compliance of the *Policy*. We note that the Research Support Fund (RSF) funding received by the University will be insufficient to support the additional costs that will be created by this *Policy*. The draft *Policy* is also unclear as to whether institutions or the Tri-Agencies will have a role in monitoring and assessing RDMPs, which may also have implications for resources. A phased in approach for the implementation of the *Policy* is important in determining how the *Policy* can be operationalized both within and, potentially, across institutions, including securing expertise to develop and maintain the necessary supports and infrastructure.

Privacy and confidentiality

We believe the *Policy* needs to be clearer that researchers and institutions should adhere to their obligations in terms of privacy and confidentiality for research data, and if there is a conflict between the *Policy* and other obligations, privacy and confidentiality obligations should take precedence.

Indigenous research

RDM for Indigenous research data, if implemented appropriately, may open new avenues for collaborations between academic and Indigenous communities, but it is important that the *Policy* considers the implications for Indigenous research and knowledge sharing. Given the current context of the Truth and Reconciliation Commission, we believe the *Policy* should be explicit that where research involves Indigenous peoples in any way, additional data management and data governance requirements must be met. Any Tri-Agency-funded researchers working with data of any kind or type that involves Indigenous peoples should be required to develop a data management plan that is jointly developed by the community or communities representing those researched. Increasingly, individual First Nations are also developing or have developed their own requirements for research ethics protocols and/or research data governance models that must be respected, which include First Nations management of and/or participation in the governance of data created about them.

Definition of research data

The draft *Policy* defines research data as “...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.”

This definition was intended to be broad to address the breadth of scholarship. We believe it would be most helpful if the *Policy* allowed for the definition of data to be specified by researchers as appropriate to discipline norms and standards.

In terms of the definition of data that are to be made accessible and archived as related to the data collected throughout the research life cycle, we recommend the definition be narrowed to data that have been publicly reported.

Given the importance of the definition of data, we believe the definition should be in the *Policy* itself; if it is referenced from an existing standard (such as the CASRAI standards in the draft *Policy*) that should be included as an appendix rather than as a link to a website that may change with time.

Recognized repository services

The draft *Policy* highlights the role of “recognized repository services or other platforms that securely preserve, curate and provide continued access to research data.” Clarity is needed as to the criteria or principles that define recognized repository services and how they would be recognized as appropriate- for example, ensuring that they meet requirements for data security, access, preservation and stability. Such criteria should appropriately be set out within the *Policy* itself and not in associated Guidelines or Frequently Asked Questions.

It is important that the *Policy* not highlight a particular tool for data management. Different scholarly communities and institutions may have their own standards and prefer to share a variety of tools and resources. For example, in Canada, the Assembly of First Nations has supported the creation of a non-profit which addresses these issues and provides training to researchers in the concept of OCAP (Ownership, Control, Access and Protection): The First Nations Information Governance Centre/Centre de gouvernance de l’information des Premières Nations (fnigc.ca). Its mission statement clearly articulates the need for distinct data management plans that incorporate First Nations data governance when data involving First Nations is to be released to the public. This is just one possible model of a type of digital repository that could be appropriate for some types of data.

Data archiving timespan

The minimum length of time data must be retained and made accessible is not specified, and this would be critical in assessing institutional and/or researcher community infrastructure and operational requirements.

Requirements for and evaluation of RDMPs

The *Policy* notes that “The agencies encourage grant applicants to complete data management plans (DMPs) as an essential step in research project design.” We assume that the language is broad rather than specific in requiring an RDMP, as a particular agency may specify requirements within a particular grant process or specific to a scholarly discipline. If this will be the case, it would be helpful for clarity in this section, as well as an implementation date.

There is no specification in the draft *Policy* as to what constitutes an acceptable RDMP. The *Policy* is unclear about whether the Tri-Agencies or institutions will have responsibility for monitoring and assessment and, importantly, how an RDMP might be assessed in a grant application phase. The latter is

particularly important given that RDMPs may involve a change in current practice for many scholarly disciplines.

Thank you for the opportunity to provide feedback on the draft *Policy*. We fully support this initiative and look forward to continuing to work with the Tri-Agencies on its implementation. We would appreciate additional consultation as the Policy and associated agency operational processes are further developed.