What We Heard Report: Engagement on the data deposit requirement of the Tri-Agency Research Data Management Policy

Social Sciences and Humanities Research Council of Canada
Conseil de recherches en sciences humaines du Canada
Canada

© His Majesty the King in Right of Canada, as represented by the Minister of Industry, 2025

Catalogue Number CR22-136/2025E-PDF
ISBN 978-0-660-77875-4

What We Heard Report: Engagement on the data deposit requirement of the Tri-Agency Research Data Management Policy

Contents

1. Executive summary 1
2. Background 4
3. Methods of engagement 5
4. What we heard 5
4.1 Clarity of the policy text 6
4.2 Timeline for implementation 7
4.3 Capacity and needs 8
4.4 Data repositories 9
4.5 Data curation 10
4.6 Data sharing 11
4.7 Sensitive data 12
4.8 Indigenous data sovereignty 13
4.9 Big data 15
4.10 Monitoring and compliance 15
4.11 Other considerations 16
5. Conclusion 17
6. Next steps 18

1. Executive summary

This report summarizes feedback received 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) (the agencies) regarding the implementation of the data deposit requirement of the Tri-Agency Research Data Management Policy. Engagement sessions took place between July 2023 and January 2025. Their purpose was to gather information on various perspectives of interested parties; developments in research data management (RDM) policy and practice, both nationally and internationally; the challenges, needs, capacity and readiness for data deposit in Canada; and what a workable timeline would be for implementing the requirement. The agencies employed various methods of engagement, including online and in-person meetings with interested parties and partner organizations within the research community. Written submissions were received from several participants after the meetings.

Overall, participants expressed broad support for the data deposit requirement, and agreed on its importance—and that of the RDM Policy as a whole—in advancing research data governance in Canada. It is considered necessary to align Canadian RDM policies and practices with those of other major funding agencies and research partners. Participants believe requiring data deposit and facilitating data sharing in Canada would improve the transparency and reproducibility of publicly funded research and facilitate data reuse, thereby increasing efficiency in Canadian research. In terms of the tri-agency data deposit requirement, input from the community highlighted key insights, considerations and needs in the Canadian research ecosystem at large—including granting agencies, researchers, service providers, institutions (libraries, research offices) and partner organizations. In general, the community expressed uncertainty about its readiness for the requirement but agreed on the importance of depositing and sharing Canadian research data.

1. **Clarity of the policy text.** Participants frequently expressed the need for clarification about aspects of the data depoit requirement: the requirement's objectives for reproducibility and data reuse; phrasing around expectations of sharing as distinct from depositing data; ethical, legal and commercial obligations; definitions of terms such as "repository," "curation" and "appropriate access"; and the applicability of participants' consent and researchers' obligations in light of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2).
2. **Timeline for implementation.** Participants had mixed opinions about a workable timeline for implementing the data deposit requirement. Some pointed to a lack of readiness among interested parties and worried about scaling services and infrastructure to meet anticipated demand in the short term (two to three years). Others argued for rapid implementation to align with international standards and catalyze institutional readiness. The advantages of a phased approach to accommodate varying levels of preparedness across research communities and institutions were explored. Overall, perceptions of readiness for data deposit differed among members of the community, but there was a general appreciation of the need to advance data deposit and data sharing in Canada.
3. **Capacity and needs.** The case was made that Canada benefits from a robust foundation of expertise, infrastructure and services to support RDM and data deposit. However, additional resourcing will be essential to scale services and capacity with the anticipated increase in demand for RDM services when data deposit becomes mandatory—and to address specific challenges associated with managing and depositing sensitive data, Indigenous data and big data. Many institutions, especially smaller ones, lack capacity and resources for RDM, which creates challenges for broadly implementing data deposit. Strong concerns were voiced about the scalability of RDM services, namely data curation services and researcher training, as well as an increase in workload for researchers and support staff, without additional resourcing.
4. **Data repositories.** Participants generally agreed that sufficient digital infrastructure currently exists for most Canadian researchers to deposit nonsensitive data into data repositories (as opposed to active data storage). However, RDM experts and service providers pointed to the importance of being able to find and access data in these repositories, rather than just using them for data retention purposes. Participants requested guidance on selecting suitable repositories. Some would

prefer a list of recommended options while others preferred to know which desirable characteristics to seek out when choosing a data repository.

5. **Data curation.** Discussions touched on the importance of minimum curation standards for deposited data to be as FAIR (findable, accessible, interoperable, reusable) as possible. It was suggested that the level of curation should depend on the data's nature and intended use. Curation was generally seen as a shared responsibility involving researchers, institutions and repositories, and the critical need for additional institutional, regional and system-wide funding to support curation services was repeatedly stressed, since curation is necessary to achieve FAIR data.
6. **Data sharing.** Participants expressed broad support for the idea that deposited data should be "as open as possible, as closed as necessary." Research security and access control were seen as critically important to protect sensitive data, Indigenous data, and data with intellectual property considerations. Open metadata was also deemed important for findability, particularly with restricted-access data.
7. **Sensitive data.** The current lack of a national and institutional infrastructure for depositing sensitive data and controlling access to it was a recurring theme throughout the engagement. Institutional data custodianship and retention policies and workflows for sensitive data are also insufficiently defined and implemented at most institutions. Participants stressed the need to clarify the definition of sensitive data, calling for guidance on sensitive data management, retention and access while repository options are being developed. Participants raised concerns about differing disciplinary standards, the need to consult research ethics boards about ethical data deposit and access, and the need for flexibility regarding commercially sensitive data.
8. **Indigenous data sovereignty.** The data deposit requirement aligns with Indigenous peoples' right to self-determination. It recognizes that data related to research by and with First Nations, Métis or Inuit communities whose traditional and ancestral territories are located within Canada must be managed in accordance with data governance principles developed and approved by these communities, and on the basis of free, prior and informed consent. Participants strongly recommended that agencies co-develop guidelines and data deposit options with Indigenous rights-holders and governance organizations, while explicitly recognizing that Indigenous governments are sovereign entities with authority over their own data.
9. **Big data.** Participants recommended clarifying the definition of "big data," and discussed the need to collaborate with national host sites for advanced research computing to clarify how the data deposit requirement would apply to very large datasets.
10. **Monitoring and compliance.** Participants sought clarification on whether compliance monitoring would apply to data deposit alone or also to data reusability in terms of the FAIR principles. Several participants argued that the agencies should be responsible for compliance monitoring, potentially by conducting random checks and using persistent identifiers (PIDs). Suggestions included using data management plans (DMPs) as a starting point for monitoring, and involving compliance with end-of-grant reports and future funding applications. Harmonizing compliance with open-access policies was also suggested.
11. **Other considerations.** Additional considerations included: making sure the data deposit requirement would not deepen existing inequities among researchers and institutions; clarifying the relationship

between depositing, retaining and preserving data; the challenges of long-term custodianship of restricted data; the ethics of data reuse; and the complexities of code deposit.

2. Background

In March 2021, 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) adopted the Tri-Agency Research Data Management Policy (the policy), with the objective of supporting Canadian research excellence through sound research data management (RDM) and data stewardship practices.

The policy consists of three requirements. First, each postsecondary institution and research hospital eligible to administer CIHR, NSERC or SSHRC funds is required to create an institutional RDM strategy. Second, for certain funding opportunities, applicants must submit a data management plan (DMP) to the appropriate agency at the time of application. Third, grant recipients are required to deposit into a digital repository all digital research data, metadata and code that directly support the research conclusions in journal publications and preprints that arise from agency-supported research.

The agencies committed to implementing the policy incrementally, based on input from engagement with the research community and other interested parties, and in step with the continuing development of research data practices and capacity in Canada and internationally. The policy's final requirement, data deposit, is to be implemented after reviewing institutional strategies and assessing the readiness of the Canadian research community.

Insights from a recent cross-sectional study mapping over 200 Canadian institutional RDM strategies are available here.[1] This analysis found that RDM strategies indicate significant progress among institutions in aligning research activities with the Tri-Agency RDM Policy. However, more work is needed with regard to consultation gaps, resource constraints, DMP and data deposit support and tools, to ensure all institutions respect Indigenous data sovereignty. In particular, many institutions reported capacity issues in terms of scaling infrastructure and services to meet the potential increase in demand for RDM services once the tri-agency DMP and data deposit requirement are implemented.

From July 2023 to January 2025, the agencies engaged broadly with members of the research community in Canada and internationally to better understand:

- various perspectives of interested parties affected by the data deposit requirement;
- developments in RDM policy and practice, both nationally and internationally;
- the challenges, needs, capacity and readiness for data deposit in Canada;
- what a workable timeline would be for implementing the requirement.

[1] Ripp C, Armond AC, Albert M, Apavaloae A, Cooper A, Costanzo L, Dearborn D, Franks A, Khan S, Lartey E, MacKinnon K, Moher D, Perpalaj F, Roche DG, Ramos J, Steeleworthy M, Wang M, Cobey K (2024), Mapping Canadian institutional research data management strategies: A cross-sectional study. *OSF preprints* https://doi.org/10.31219/osf.io/tnkp9

3. Methods of engagement

The agencies led and participated in numerous engagement meetings with partner organizations and members of the research community, both online and in-person. Partners included, among others, the Digital Research Alliance of Canada's RDM leadership and expert groups, Canadian service providers (Borealis, the Federated Research Data Repository, SciNet, Calcul Québec), the Office of the Chief Science Advisor of Canada, the Canadian Research Data Centre Network, the Secretariat on Responsible Conduct of Research, members of tri-agency Indigenous advisory structures, national and international granting agencies, science-based departments and agencies of the Government of Canada, intergovernmental agencies, and organizations from the not-for-profit and private sectors. Members of the research community included RDM specialists, librarians, researchers, research administrators, members of the Network of Indigenous Research Administrators, research security and IT specialists, research ethics board members, and research legal counsels. In addition to verbal input received during this engagement, the agencies also received written submissions from several interested parties. CIHR University Delegates and NSERC and SSHRC Leaders also gave their feedback and input.

Both broader engagement sessions and meetings with individuals or small groups of partners and interested parties were held. Between July and September 2023, four virtual meetings took place with the RDM Network of Experts Council of Chairs of the Digital Research Alliance of Canada. Two in-person engagement sessions were held in spring 2024, one at the University of Calgary after the 2024 Canadian Conference on Research Administration (46 participants), and another in Halifax during the 2024 Annual Conference of the International Association for Social Science Information Service and Technology (roughly 50 participants). In 2024, three virtual information sharing sessions and town halls were organized, upon request by universities. Three virtual engagement sessions also took place in January 2025, two in English and one in French. These sessions were widely advertised, and over 350 participants from more than 200 institutions across Canada attended. Participants were invited to provide input during a question and discussion period, and via anonymous answers to several online polling questions. Feedback from participants and groups quoted in this report has been de-identified.

4. What we heard

Below is an overview of feedback and input the agencies received on key issues in implementing the data deposit requirement. While this report is intended to be comprehensive, it might not reflect all the considerations that arose. Nevertheless, the agencies took note of all comments and will integrate them as much as possible going forward.

Overall, participants voiced strong support for data deposit, and saw the requirement and the RDM policy as highly beneficial for research data governance and RDM practices in Canada. Feedback from the community shed light on critical insights, needs and considerations in terms of the Canadian research ecosystem, including granting agencies, researchers, service providers, institutions and other organizations that play a role in Canadian RDM. This input addressed issues directly related to implementing the tri-agency data deposit requirement, including considerations beyond the RDM policy's scope or the agencies' remit but still relevant to the wider community and partner organizations.

4.1 Clarity of the policy text

Participants repeatedly called for clarification of the objectives and advantages of the data deposit requirement. The text should expand on how data deposit can serve different purposes, such as validating research results (i.e., transparency and reproducibility) to increase trust in published findings, and how data sharing can increase reuse and efficiency in Canadian research.

Several participants, including RDM experts, also pointed out areas of the policy text that ought to be made more clear. For example, after the sentence about grant recipients being "required to deposit all digital research data, metadata and code that directly support the research conclusions in journal publications and pre-prints," the statement about recipients not being required to provide access to data created confusion around whether deposited data must be publicly shared: "Grant recipients are not required to share their data. However, the agencies expect researchers to provide appropriate access to the data where ethical, cultural, legal and commercial requirements allow, and in accordance with the findable, accessible, interoperable, reusable (FAIR) principles and the standards of their disciplines." This wording could be misinterpreted by researchers and other interested parties, possibly undermining the policy's intent. A stronger affirmation, as in the National Institutes of Health statement's "researchers are expected to maximize the appropriate sharing of scientific data," was recommended, to declare the value of data sharing while allowing for the necessary restrictions applicable to sensitive data, Indigenous data and public-private data, for example.

Participants said ambiguities around "data sharing" and "data deposit" have generated confusion within the community. The agencies should clarify what they mean by "sharing," as well as what "ethical, legal and commercial obligations" apply to making data available to others for reuse. Questions arose around whether the requirement's default was public sharing or would be sufficiently met by depositing data in a repository without making it publicly available. Recommendations were made about distinguishing between requirements for sharing data as opposed to sharing metadata, with the suggestion that some metadata should always be openly available regardless of accessibility to its primary data, to ensure that research data are stored in appropriate locations and remain findable/discoverable regardless of sharing obligations.

Many participants recommended defining key terminology in the policy text, saying the lack of clear definitions of these terms would pose challenges for implementing the requirement at the institutional level. A comprehensive glossary of terms would help ensure all interested parties have a common understanding of the terminology used in the policy. Participants pointed out that many terms, including "repository," "safe storage," "preservation" and "curation," could be variously interpreted, while others, such as "appropriate access," "sharing" and "disciplinary standards," were ambiguous. Also, it is unclear how disciplinary standards are determined, who sets them, and what happens if disciplinary norms disagree with the policy. It was noted that the lack of clarity creates uncertainty and the potential for conflict in the implementation process; the policy should address how differences between disciplinary standards would be reconciled.

The agencies heard that the policy should specify what constitutes "minimal compliance" with the data deposit requirement. Recommendations included providing examples and scenarios to help guide

researchers in understanding the practical implications of the requirement; specific guidance could help avoid inconsistent implementation across disciplines and institutions.

Some participants suggested the policy should treat “research data, metadata and code” as distinct categories, each with its own set of requirements. This separation would help simplify policy implementation by allowing these elements to be treated differently, for example by enabling metadata to be shared publicly to enhance findability even when data or code are restricted.

The agencies heard that the policy needs to address the relationship between the data deposit requirement and the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2)—specifically, how the data deposit requirement aligns with the new TCPS2 article on broad consent, where participants must give separate consent for data deposit intended for publishing results and for future research. The policy should clarify if and how researchers who deposit data intended to support published results (e.g., in the case of specific consent) will limit data sharing and reuse for study participants who did not consent to their data being used for future, unspecified research (i.e., broad consent).

> Guidance around the alignment between this policy and TCPS 2 (2022) would benefit researchers who conduct research on human participants.
>
> (RDM Expert Group)

4.2 Timeline for implementation

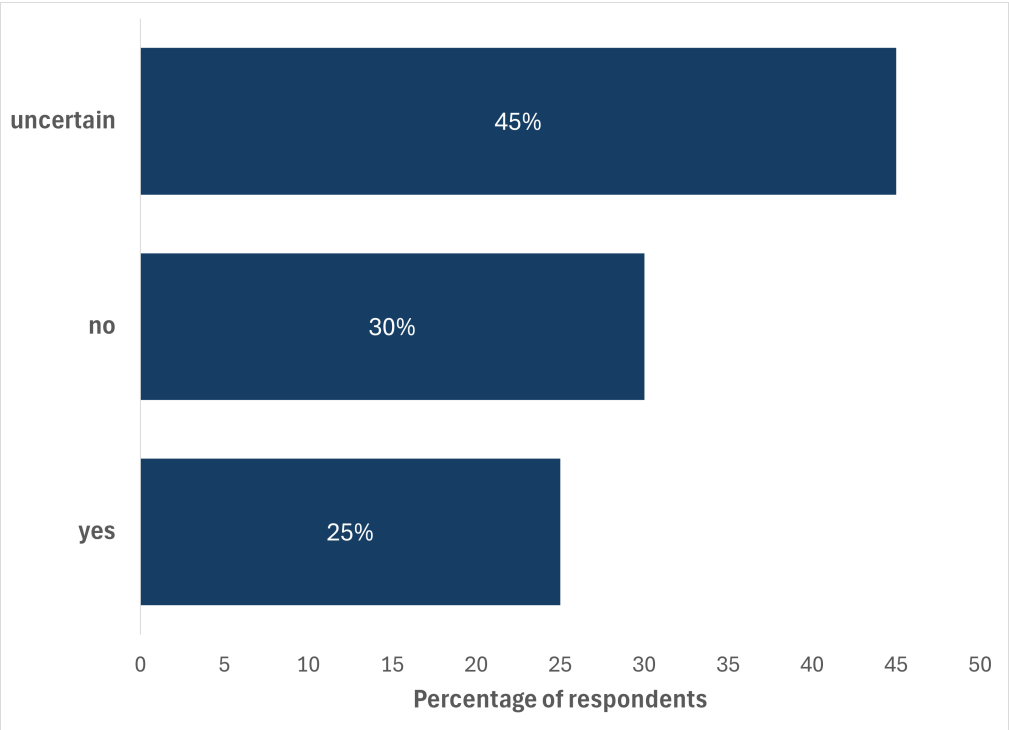
Participants expressed different perspectives on an appropriate timeline for rolling out the data deposit requirement (Figure 1). Some voiced concern that a short timeline might not allow sufficient time for researchers, institutions and repositories to prepare for the requirement’s coming into effect, particularly given the complexities involved (e.g., considerations surrounding sensitive data, Indigenous data, big data, curation and support). Conversely, others emphasized that Canada is lagging behind other major research funders in requiring data deposit and facilitating sharing, and that the Canadian research community could never achieve the level of readiness to allow for flawless implementation. The importance of implementing the requirement sooner rather than later to generate momentum for increased readiness within the community was emphasized. Some participants believed announcing an implementation date would help catalyze institutional readiness; preparation for readiness would improve once a date was set, since institutional leadership is often hesitant to commit funds to any endeavour not yet operationalized. It was noted that transparency about the implementation timeline and the agencies’ plan is needed to build trust and support among the community.

Many participants suggested a phased approach to implementation to manage these complexities and the varying readiness levels of different research communities and institutions. A phased approach would allow for a gradual integration of the requirement, starting with simpler cases and progressing to more complex ones, such as sensitive data or very large datasets. However, concerns were also voiced that a phased approach, similar to the Data Management Plan (DMP) roll-out, might take too long, potentially significantly delaying the full implementation of the data deposit requirement. The importance of DMPs in helping researchers plan for data deposit was also emphasized, given that DMPs

are meant to detail where and how research data will be managed, stored, deposited and shared, where appropriate.

> Not all institutions are ready, but the bar is low enough that they should be able to do so sufficiently once a date is formally set. Canada seems to be lagging on this front, compared to the US, Europe, Japan and other countries. Important to just set a date and give institutions a real target.
>
> (Engagement participant)

Figure 1. This figure shows the percentage of respondents (N = 138) who replied "yes," "no" or "uncertain" to the question, "Given the lag between funding and publication, do you think the Canadian research community will be ready to comply with a data deposit requirement applicable to grants awarded after January 1, 2026?" The question was asked in January 2025 through a live poll, during virtual engagement sessions on implementing the data deposit requirement.



4.3 Capacity and needs

Many participants shared concerns about the lack of capacity, resources and infrastructure to support the data deposit requirement. Representatives of institutions, particularly smaller ones, such as colleges and cégeps, expressed worry about their ability to meet the demands of the data deposit requirement, since they have fewer resources, less access to expertise, and more limited capacity for implementing complex policies than large institutions have. Participants mentioned that, despite their published RDM strategies, institutions need to develop long-term strategies and implementation plans for supporting data management and deposit.

The demand for funding to support data curation services and training for researchers and support staff was repeatedly mentioned throughout the engagement: without dedicated resources for training and support, participants said many institutions would struggle to implement the requirement. Some participants emphasized the need for executive sponsorship and leadership within institutions to prioritize and support the implementation of RDM practices.

Many participants said researchers and institutions are struggling with time constraints: the time required for proper data management, curation and deposit is a major concern. The agencies heard that RDM is often viewed as an additional task rather than an integrated part of the research process and workflow. Research administrators and librarians are already overburdened, and the data deposit requirement could add a significant new task without new resources.

> There needs to be more funding for organizations to support this initiative, and more education for researchers, which also requires funding. I think expecting institutions to use RSF [Research Support Fund] and Research Security stipend funding is not reasonable. This funding is already spoken for.
>
> Many of us are running research teams on small grants and a fixed budget, but we're facing significant inflation in costs. The additional costs of data management will be a challenge for many of us.
>
> (Engagement participants)

4.4 Data repositories

Conversations with service providers, researchers and RDM experts underscored the breadth and robustness of existing infrastructure and services for data deposit in Canada. The Federated Research Data Repository and Borealis, among others, provide researchers with a free-to-use, Canadian-hosted repository infrastructure. Data discovery is enabled by lunaris, a national, bilingual discovery service for data from research institutions across the country. While there are limitations regarding the type of data that Canadian repositories currently accept, initiatives are underway to broaden their scope and provide solutions for controlled access management of sensitive data, where deposit is possible and appropriate.

Participants emphasized the importance of a clear definition of what constitutes an adequate repository to help researchers differentiate data repositories from other data storage solutions, such as active storage. Many researchers are storing their data on institutional servers or cloud platforms, but it is unclear whether these practices meet the data deposit requirement, and this has caused confusion. Guidance on "desirable characteristics for data repositories," as provided by the National Institutes of Health and the United States Office of Science and Technology Policy, would help researchers select a suitable repository without being overly prescriptive. The point was also made that the agencies have a role to play in encouraging the use of a Canadian-developed infrastructure, particularly in the context of storing/depositing and safeguarding sensitive data.

Many participants mentioned the need to establish minimum standards for repositories, including sustainability plans, the ability to issue persistent identifiers (e.g., digital object identifiers [DOIs]), and retention periods to ensure safe storage and preservation. For example, repositories should have a

mandate and plan for data preservation; provide information to help people discover and learn about deposited data; provide potential data users with either direct access or information about access conditions; and ensure each data set has a persistent identifier (PID) that uniquely identifies the dataset and facilitates citation in publications.

The agencies heard that requiring repository certification, such as CoreTrustSeal, may be too restrictive, severely limiting options available to Canadian researchers. Many generalist repositories, including those partly or wholly funded by the Canadian federal government, are widely regarded by RDM experts as sound options for depositing data, but they do not have the resources to obtain CoreTrustSeal certification. Certification is not the only way to prove that a repository is trustworthy.

The point was made that making research data findable and publicly accessible is a primary mission of data repositories. As such, many repositories, including Canadian service providers, would be unlikely to support researchers who deposit nonsensitive data while restricting access to them (i.e., placing access controls on data that can be shared publicly).

> If a dataset will never be accessible under any circumstance, it is worth asking whether there is any utility at all in making it findable. Repository policies may limit those types of deposits.
>
> (RDM Expert Group)
>
> We know about Borealis and FRDR but trying to figure out all the disciplinary/subject specific repositories is challenging, and we need to be able to give guidance on those.
>
> (Research administrator)

4.5 Data curation

Data curation is the process of organizing, managing and maintaining data to ensure its quality, usability and long-term accessibility. It should be done by the data generator(s) and supported by RDM specialists. Varying views were expressed regarding the level of curation required before datasets are deposited. The amount of curation needed should depend on the type of data, its intended use and the length of time it needs to be stored. Some participants argued that setting a minimum standard for data curation would help ensure the long-term access, discovery, reuse and preservation of data. This standard could serve as a starting point, so that datasets are not just deposited but also maximally FAIR (findable, accessible, interoperable, and reusable). Curation for data interoperability and reusability was deemed important, as was the need for metadata standards. Standardized metadata is essential to ensure deposited data can be found and reused.

Participants said many institutions, even large ones, have limited capacity to assist their researchers with data curation. Nevertheless, curation should not be seen as the sole responsibility of researchers. Professional data stewards, librarians and other information professionals have the expertise needed for proper curation, and they should be involved in curation efforts. However, many participants noted the critical importance of additional funding in the system to support curation services: repositories and institutions (libraries) need resources to ensure deposited data are properly documented and managed to promote reusability.

Many participants argued that insufficiently curated data would result in many deposited datasets not being usable or FAIR, and therefore not aligning with the spirit of the data deposit requirement. Researchers' requests for curation assistance from repositories and institutional services will grow rapidly upon implementation of the data deposit requirement, and it is likely that many deposited data sets will not undergo proper curation without significant additional funding in the system.

> A key issue is resourcing for small institutions and associated, uneven levels of resourcing for data curation even in Borealis.
>
> I would really still like to see some guidelines on curation and making data usable. NLM [US National Library of Medicine] has a Data Repository finder that includes ICPSR and Dataverse and SocSci repos. If we don't guide people, they will do the very easiest thing!
>
> (Engagement participants)

4.6 Data sharing

Many participants commented on the fact that the policy requires data deposit, but not open data. While the policy encourages openness and compliance with the FAIR principles, it does not mandate open data. Participants highlighted the need to better understand what the agencies expect in terms of data deposit: it is unclear what is needed beyond simply storing data in a digital repository. It was noted that depositing data without making public information available about the data—especially metadata and persistent identifiers (PIDs)—goes against the FAIR principle of “findability.” Several RDM experts raised concerns that data deposit could become a tick-box exercise if it only involved depositing data without making them findable, accessible or reusable.

Several participants said open data should be the default expectation for data sharing, and that justifications should be required for not providing open data. Ideally, these justifications would be included in a DMP at the beginning of a research project. More than 90% of participants favoured an “as open as possible, as closed as necessary” approach in a polling question on this topic during broad virtual engagement sessions held in January 2025. The general view was that openly available data can improve data discoverability and reuse, although many data (e.g., sensitive data) need to be restricted (and stored on secure infrastructure with designated custodianship rather than on personal hard drives or cloud storage). Requiring open metadata for restricted, sensitive data—and a clear description of the request process to access these data—would help the agencies promote FAIR principles for sensitive data in Canada, in particular by increasing their findability and accessibility.

There was broad agreement that a key aim of the requirement should be to encourage transparency and reuse of data for new research. As such, the policy should require that research data supporting published papers not just be deposited, but also available for reuse (where appropriate) and well documented to ensure they are, in fact, reusable.

Expand on how sharing/deposit can be defined for different purposes such as replication, reuse, etc. Perhaps tease out in an FAQ as data sharing for replication and for future unspecified research can have ethics implications (as in TCPS 2, Chapter 3 Section E Article 3.13).

(RDM Expert Group)

Researcher responsibility is reasonably well described but what are the institutional responsibilities? We have high levels of researcher departure in the last few years: who is responsible for ensuring data accessibility after a researcher retires or leaves the country?

(Engagement participant)

4.7 Sensitive data

Engagement participants repeatedly expressed concerns about the lack of Canadian national and institutional infrastructure for depositing sensitive data and controlling access to it. Options are also limited for depositing sensitive data outside of Canadian repositories. Under provincial and territorial legislation governing sensitive data, local and regional solutions must be developed. To help align RDM policy requirements with the realities of existing Canadian repositories, a dedicated committee or council could be useful for discussing pain points, challenges and opportunities.

Some participants said a comprehensive definition of sensitive data is required in order to give researchers clear guidelines on complying with the data deposit requirement. This definition must be broad enough to encompass disciplines' various needs while identifying different risks. Guidance on managing sensitive data is particularly lacking, including data collected from research involving humans in general, Indigenous communities in particular, and research conducted by or with industry partners. This guidance should cover issues of data security, intellectual property, access control and long-term preservation.

Concerns were also expressed about different standards and levels of acceptable risk for sensitive data between disciplines. Some participants suggested that the agencies could work with different disciplines to determine what kind of data was sensitive and agree on appropriate levels of risk. Others argued that it was ultimately up to individual fields and disciplines to define sensitive data and determine related levels of risk. The argument was also made that the agencies cannot provide exhaustive and detailed guidelines and should instead provide broad, high-level guidance, particularly on preparing data for deposit and sharing (where appropriate), including de-identification and anonymization guidelines. Discussions also touched on concerns about data security and the lack of standardization for sensitive data, since the standardized data classifications used in IT may not be applicable to data deposit.

Some participants recommended a phased implementation of data deposit for sensitive data, which would allow time for Canada to develop suitable infrastructure and practices. While deposit solutions are being developed for sensitive data, the agencies should consider the feasibility of steps such as requiring that metadata be published with a DOI, made findable through major data discovery tools such as lunaris (Canada's national data discovery service), and accompanied by a clear process for requesting access to restricted data. However, the point was also raised that data repositories, particularly Canadian ones, typically do not accept deposits of unaccompanied metadata records, given the challenges of

ensuring the permanence of, and access to, primary data stored elsewhere. It was noted that solutions would have to be developed in collaboration with service providers, including the Digital Research Alliance of Canada.

Participants noted that data de-identification is often necessary for ethical access to sensitive data. The policy should clarify that providing access to de-identified data for reuse does not contravene the policy, as long as research participants' privacy is ensured and they gave their consent. It was also pointed out that metadata (even in the absence of primary data) can pose a risk of re-identification, creating the need for guidance on assessing and mitigating that risk.

Several participants advised consulting research ethics boards about managing sensitive data, since they are well equipped to help determine requirements for ethical data deposit and access control. Potential discrepancies between the data deposit requirement and the TCPS2 broad consent mandate were also discussed (see 4.1 Clarity of the policy text). It was recommended that the policy clarify requirements to ensure that ethical and legal obligations are met when depositing data and sharing anonymized data and/or metadata on human participants.

The agencies heard that there were valid reasons for not depositing ad/or sharing certain commercially sensitive data and other types of sensitive data, and these reasons must be acknowledged and flexibility built into the policy to accommodate them. When data cannot be deposited in a recognized repository, researchers should provide an explanation and an alternative plan for data retention and data preservation, if applicable.

> A listing of viable sensitive data types and minimal protection requirements would need to be provided, both for researchers and data repositories to reference.
>
> Open metadata infrastructure for closed and restricted/sensitive data does not exist at the national level and is not integrated with existing infrastructure.
>
> (RDM specialists)

4.8 Indigenous data sovereignty

There was broad consensus among participants that Indigenous data sovereignty is essential for Indigenous self-determination, the ongoing work of reconciliation and successful researcher collaboration with Indigenous rights-holders. During engagement sessions, the agencies pointed out how the policy emphasizes the importance of Indigenous rights-holders' control over the collection, ownership, protection, use and sharing of data that concerns them, in accordance with the data governance principles they developed and approved. Participants affirmed that the policy's commitment to Indigenous data sovereignty is critical to ensure researchers do not deposit any data against the wishes of Indigenous rights-holders.

Many participants mentioned the need to co-develop guidelines and options for data deposit with Indigenous communities and governance organizations. This would involve holding discussions with Indigenous rights-holders and ensuring they have the right to decide if and under what conditions their

data might be deposited. Indigenous-led resources were identified to help researchers, research administrators, funders and other parties ensure adequate levels of engagement and co-development in research involving Indigenous rights-holders. Examples included the Indigenous Research Level of Engagement Tool (IRLET) and the Circumpolar Inuit Protocols for Equitable and Ethical Engagement, developed by the Inuit Circumpolar Council (EEE-Protocols-LR-WEB), among others. Participants also noted the importance of the principles of ownership, control, access and possession (OCAP®) as a framework for First Nations data sovereignty. However, the policy acknowledges that this model may not always respond to the needs and values of distinct First Nations, Métis and Inuit communities, collectives and organizations, and that a distinctions-based approach is needed to ensure acknowledgement and respect for the unique rights, interests and circumstances of First Nations, Métis and Inuit communities.

The agencies heard that some repositories may need to develop policies and procedures to respect Indigenous data sovereignty by supporting Indigenous communities in governing their data and protecting their rights. Participants recommended that the policy should define what constitutes a community, and ensure that sovereign Indigenous governments have decision-making authority over research and data involving their citizens. In addition, the policy should explicitly acknowledge and support the memorandums of understanding and data agreements commonly used in research by and with Indigenous Peoples to define the terms and conditions for data management, deposit and access.

The following recommendations were proposed to help the agencies promote Indigenous data governance and sovereignty: clarify the concept of Indigenous data by providing a definition in materials associated with the policy; use plain language in agency documents, DMPs and research proposals, so they can be understood by all parties concerned; make DMP templates flexible enough to adapt to communities' specific needs and distinct characteristics; require grant applicants to describe in their DMPs how data management was co-developed through meaningful engagement with Indigenous rights-holders; offer grants to encourage and enable early engagement with Indigenous rights-holders prior to developing a research proposal; and support the development of engagement guidelines for researchers who belong to the community in which they are conducting research.

Members of the tri-agency Indigenous Leadership Circle in Research and the Reference Group for the Appropriate Review of Indigenous Research stressed the importance of critical infrastructure investments in data management and storage for Indigenous communities, collectives and institutions. They also expressed the need to keep knowledge connected to the community. Noting that some concepts, such as “data” and “intellectual property” do not align with many Indigenous worldviews, they encouraged framing data deposit to emphasize the preservation, protection and transmission of traditional knowledge for future generations.

> When it comes to DMPs and data deposit, it’s nothing about us without us. If a community does not have time, capacity or interest to engage on a research project, then that research should not go ahead.
> (Indigenous research administrator)
>
> Indigenous governments are rightsholders and must be consulted on DMPs and data deposit.
> (Researcher)

4.9 Big data

The agencies heard that policy implementation should take into account considerations related to very large datasets, or “big data.” Participants said management, storage, retention and sharing practices for very large datasets arising from projects using advanced research computing (ARC) resources differ significantly from those of smaller datasets, which can readily be deposited in data repositories.

Several participants noted that many data repositories lack the infrastructure needed to support big data. Studies conducted using ARC generate large volumes of data that need specialized handling and storage capabilities many current repositories may lack, while ARC systems do have such capabilities. In addition, there are significant challenges associated with transferring very large datasets to data repositories (such as datasets in the hundreds of terabyte or petabyte range), even if the repositories have the infrastructure to handle them. As a result, the deposit requirement should allow for big data to remain within the systems designed for them, even though big data sites, in general, have yet to fully align with FAIR principles (e.g., data are neither registered with PIDs nor discoverable; common metadata standards are not used; storage is either not guaranteed or requires re-applying for allocation).

The agencies were advised that further engagement and close collaboration with the National Host Sites for ARC was needed to ensure that the data deposit requirement makes sense and is practicable as it applies to big data. Areas for further discussion involve public metadata sharing and processes for direct access and access control to data maintained by the National Host Sites.

> From the perspective of the ARC/HPC [high performance computing] community (sysadmins, analysts, managers, and the researchers themselves), if they are already using the national sites, we have the stewardship, access and control of that data. The data resides on our storage systems. There is no need for a "deposit to a repository" step. What we need is a process by which we can conform that data to RDM best practices, with the consent of the researchers.
>
> (ARC storage analyst)

4.10 Monitoring and compliance

Some participants mentioned clarifying the objectives of monitoring researchers' compliance with the data deposit requirement. It would be helpful to clarify the policy about what compliance the agencies intend to ensure: simply that datasets are deposited, or also their alignment with FAIR principles to enable research verification and data reuse? Many participants argued that the agencies should consider how to ensure that deposited data are (re)usable: compliance monitoring should assess whether data is not only present, but also well-documented and reusable. However, it was acknowledged that, since the agencies cannot police every dataset, compliance monitoring must focus on a manageable and effective set of criteria such as essential metadata, clear access procedures for restricted data, and nonproprietary file formats, among others. The agencies could implement random checks of deposited data, use PIDs to connect published papers with datasets and agency grants, and remind researchers of their obligations when needed.

Several participants pointed to DMPs as a starting point for monitoring and compliance: DMPs could detail anticipated outputs and be updated over time to reflect changes in the study. DMPs could also be used to assess researchers' fulfilment of their commitment to deposit and publicly share data, if appropriate, or share metadata and provide controlled access to restricted data. Additionally, data deposit compliance could be included in end-of-grant reports and/or subsequent applications for agency funding opportunities. Grant reports and applications could use the PIDs of deposited datasets and code associated with agency-funded research to improve discoverability.

Several participants suggested linking compliance monitoring to the Tri-Agency Open Access Policy on Publications and harmonizing compliance mandates for open access and data deposit. Responsible research assessment should be central to compliance mechanisms and incentives for data deposit and sharing, as recommended in the San Francisco Declaration on Research Assessment.

Concern was expressed about how compliance will be monitored for data deposited outside of repositories—for example, sensitive data stored on institutional servers or in cloud storage. Participants said the agencies should clarify how they would verify the deposit of restricted data held on closed platforms, since such verification could pose challenges. A compliance monitoring system for restricted datasets must ensure that deposited data and metadata remain findable and data access is easily requested and properly managed.

> Could eligibility for the next grant be contingent upon demonstrating compliance the time before?
>
> Incentives for sharing data can include: research assessment reforms to value a wider variety of research outputs (datasets, software, societal contributions, etc.), subsequent access to more funding, and promotion of research outputs.
>
> (Engagement participants)

4.11 Other considerations

Fairness and equity in research. Many participants said the data deposit requirement should not generate or deepen existing inequities among researchers, research groups, or institutions. Implementation processes should take into account the fact that certain groups would have more trouble managing the added burden, and accordingly make resources and support available.

Data retention and preservation. Several participants recommended clarifying the connection between data retention, preservation and deposit. Data retention and preservation policies among digital repositories vary, with many different actions to be taken depending on the type of research data being deposited and the level of preservation granularity. The agencies could consider working with the community to establish common retention recommendations and/or requirements for deposited data. Participants mentioned storing data on a server and without deleting them as distinct from implementing digital preservation practices that ensure content remains accessible over time. Consideration should also be given to how long datasets and code must be kept for verification and reproducibility purposes, as opposed to being kept for reuse and future research.

Data custodianship. The agencies heard that, at present, there are no guidelines or principles for determining who should be responsible for managing access control to restricted, sensitive data in the long term. The issue of licensing and copyright for all data was also raised: unless data are assigned an open license, repositories and institutions will have trouble managing data over time without researcher involvement. Participants said more information is required about the definition of "data custodianship," why it is desired, and what custodianship gives institutions, organizations and their staff in terms of ownership, responsibilities, action and resource requirements. It was noted that all (or most) repositories would be unlikely to consent to assuming custodianship of data, nor would it be readily feasible for institutions to identify and assume custodianship of datasets scattered across current repositories. Regardless of how data custodianship and its related responsibilities are defined, these could be considered and documented at the outset of a project, in a DMP. Data custodianship could define the length of time a researcher is required to share their data and/or the maximum amount of time a repository agrees to store deposited data.

Ethics of data reuse. Some participants pointed out the need for a policy component on data (re)users and their responsibilities, including ethics and proper credit attribution. It was suggested that the policy should consider the ethical and legal requirements of reusing research data.

Code deposit. The agencies heard that the degree of resources and expertise required to enable FAIR code deposit and sharing varies substantially, not least because code used for research purposes ranges from simple analysis scripts to complex software packages and libraries that require specific computing environments to execute. The degree to which researchers use code for their analyses, as opposed to software with a graphical user interface, also varies greatly. There is strong agreement in the Canadian curation community that the knowledge, skills and capacity required for code curation significantly lag those applicable to data curation, which could pose a challenge for operationalizing a requirement for FAIR deposit of code alongside data. Also, few existing data repositories are suitable for code deposit: distributed version control platforms such as GitHub are commonly used for sharing code, but they lack key functionality of data repositories in terms of accessibility, interoperability and preservation.

> Best practices for code curation have not reached the breadth and maturity of those for data. More training and collective knowledge building is required to meet the expected need. Commercial services like GitHub are an incredibly valuable portion of the research software and code development landscape; it's also important to note that their mandates aren't entirely aligned with not-for-profit repositories and their policies, offerings, and pricing may be altered at any point to meet corporate needs.
>
> RDM Expert Group

5. Conclusion

This engagement revealed critically important considerations for the data deposit requirement of the Tri-Agency RDM Policy. It included the **various perspectives of interested parties**, including researchers, librarians, RDM and IT specialists, research administrators, representatives of Indigenous groups and various organizations. Common themes included the desire for clear and practical guidance, the

importance of support for researchers, concerns about increased workloads and costs in institutions, and the imperative to respect Indigenous data sovereignty.

In terms of developments in **RDM policy and practice both nationally and internationally**, participants noted that Canada appeared to be behind other major research funders in terms of requiring data deposit and sharing. Discussions revealed that it is important to align with international best practices such as the FAIR principles, and participants mostly agreed with the need for research data to be "as open as possible, as closed as necessary." Discussions also touched on the evolving landscape of data repositories and the challenges of managing, depositing and sharing certain data types, especially sensitive data and big data.

In terms of **challenges, needs, capacity and readiness**, the community stressed the importance of additional resources to support RDM services and infrastructure, particularly for smaller institutions. Frequently cited needs for successful implementation were: definitions of key terms; guidance on sensitive, commercial and Indigenous data; and funding for critical activities such as curation and training.

Opinions differed strongly on a workable **timeline for implementing the data deposit requirement**. Some participants advocated for swifter adoption to drive change, while others cautioned that services and infrastructure were not yet fully ready. Support was expressed for a phased implementation approach to accommodate varying readiness levels and the complexities associated with different types of research data. Overall, the engagement confirmed the necessity for greater clarity, more resources and ongoing dialogue to ensure the successful and equitable implementation of the data deposit requirement.

6. Next steps

The agencies will draft an implementation plan for the data deposit requirement, with added clarifications based on input from this engagement. The plan will then be shared with the research community for feedback.

Contact us:

CIHR: ResearchData-Donneesderecherche@cihr-irsc.gc.ca
NSERC: ResearchData-Donneesderecherche@nserc-crsng.gc.ca
SSHRC: ResearchData-Donneesderecherche@sshrc-crsh.gc.ca

For information on the Tri-Agency Research Data Management Policy, visit:
https://science.gc.ca/site/science/en/interagency-research-funding/policies-and-guidelines/research-data-management/tri-agency-research-data-management-policy