

Tri-Agency Statement of Principles on Digital Data Management

1. Preamble

The [Canadian Institutes of Health Research \(CIHR\)](#), the [Natural Sciences and Engineering Research Council of Canada \(NSERC\)](#), and the [Social Sciences and Humanities Research Council of Canada \(SSHRC\)](#) (the agencies) are federal granting agencies that promote and support research, research training, knowledge transfer and innovation within Canada.

As publicly funded organizations, the agencies are strong advocates for making the results of the research they fund as accessible as possible. In promoting access to research results, they aspire to advance knowledge, avoid research duplication and encourage reuse, maximize research benefits to Canadians and showcase the accomplishments of Canadian researchers. These aspirations align with the Government of Canada's commitment to open science, as described in [Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation](#) (2014).

Research data include observations about the world that are used as primary sources to support scientific and technical inquiry, scholarship and research-creation, and as evidence in the research process.^{[Footnote 1](#)} Research data are gathered through a variety of methods, including experimentation, analysis, sampling and repurposing of existing data. They are increasingly produced or translated into digital formats. When properly managed and responsibly shared, these digital resources enable researchers to ask new questions, pursue novel research programs, test alternative hypotheses, deploy innovative methodologies and collaborate across geographic and disciplinary boundaries. The ability to store, access, reuse and build upon digital research data has become critical to the advancement of science and scholarship, supports innovative solutions to economic and social challenges, and holds tremendous potential for Canada's productivity, competitiveness and quality of life.

Governments and research funders across the globe are becoming increasingly aware of the value of digital research data, the importance of fostering reuse of digital research data and the need for policies to enable excellence in data stewardship. Canada has

joined many other countries at the forefront of this movement, as shown in its support for the Organisation for Economic Co-operation and Development's [Declaration on Access to Research Data from Public Funding](#) (2004); its commitment to the [Open Government Declaration](#) (2011); and its approval of the [G8 Science Ministers Statement](#) (2013).

The Government of Canada's [Action Plan on Open Government](#) (2014) aims to maximize access to the results of federally funded research, to encourage greater collaboration and engagement with the scientific community, the private sector and the public. The action plan includes a commitment to adopt policies to support effective data stewardship.

The agencies believe that research data collected with the use of public funds belong, to the fullest extent possible, in the public domain and available for reuse by others. They also strongly support the creation of a robust and efficient environment for data stewardship in Canada and internationally. They have encouraged data stewardship through SSHRC's [Research Data Archiving Policy](#) (1990), and data sharing provisions for CIHR grant holders in the [Tri-Agency Open Access Policy on Publications](#) (2015). They will continue to promote excellence in data management practices within the Canadian research community.

This statement of principles outlines the agencies' overarching expectations regarding research data management, and the responsibilities of researchers, research communities, research institutions and research funders in meeting these expectations.

2. Objective

The objective of this statement of principles is to promote excellence in digital data management practices and data stewardship in agency-funded research. It complements and builds upon existing agency policies, and serves as a guide to assist researchers, research communities and research institutions in adhering to the agencies' current and future research data management requirements.

3. Expectations

Data Management Planning

Data management planning is necessary at all stages of the research project lifecycle, from design and inception to completion.

Data management plans are key elements of the data management process. They describe how data are collected, formatted, preserved and shared, as well as how existing datasets will be used and what new data will be created. They also assist researchers in determining the costs, benefits and challenges of managing data. They should be developed using standardized tools, where available.

Constraints and obligations

Research data must be managed in agreement with all commercial, legal and ethical obligations.

Data management should be performed in accordance with the requirements of the [*Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – 2nd edition*](#). This statement provides guidance on data management aspects of research involving humans, such as consent, privacy and confidentiality, indigenous people's rights, secondary use of data and data linkage. Data management should also be performed in accordance with the requirements of the [*Tri-Agency Framework: Responsible Conduct of Research*](#).

Adherence to Standards

Data should be managed in accordance with the most appropriate and relevant standards and best practices, while recognizing that these are in a state of rapid evolution.

Collection and Storage

Data should be collected and stored throughout the research project using software and formats that ensure secure storage, and enable preservation of and access to the data well beyond the duration of the research project.

Metadata

All research data should be accompanied by metadata that accord with international and disciplinary best practices to enable future users to access, understand and reuse the data.

Quality metadata are essential for making research data findable, and for the systems that use or mine the data. Standards are diverse and vary across disciplines, but metadata generally state who created the data and when, and include information on

how the data were created, their quality, accuracy and precision, as well as other features necessary to enable understanding and reuse. When possible, common metadata standards should be adhered to.

Preservation, Retention and Sharing

Research data resulting from agency funding should normally be preserved in a publicly accessible, secure and curated repository or other platform for discovery and reuse by others. To determine whether data should be shared and preserved, researchers should consider the data needed to validate research findings and results, and support replication and reuse. They should look at the potential benefits that sharing the data will have for their own or other fields of research, and for society at large. Researchers should also consider whether any ethical, legal or commercial obligations prohibit sharing or preserving the data, and whether any of the data need to be de-identified or made available with restricted access.

Decisions regarding preservation, sharing and retention periods for data should be made in accordance with international and disciplinary best practices and relevant policies, seeking expert guidance where necessary. The rationale for data preservation, sharing and retention is normally defined in a data management plan.

Timeliness

Data should be shared as early as possible in the research process when they are considered to be informative and of appropriate quality.

Data release can be staged as research progresses, starting with metadata. Data supporting publications should be shared by the publication date, and where possible, should be linked to the publications. A defined period of exclusive use of data for primary research is reasonable in some cases.

Acknowledgement and Citation

Data are significant and legitimate products of research and must be recognized as such.

All users of research data should acknowledge – through citation and other practices or standards relevant to their disciplines – the sources of the data they are using, and respect the terms and conditions under which these data were accessed. Researchers

who responsibly and effectively share their data should be recognized by funders, their academic institutions and users benefiting from the reuse of the data.

Efficient and Cost Effective

Data management should be efficient and cost effective. All data need to be managed, but not all data need to be shared or preserved – costs and benefits of doing so should be considered in the data management planning process.

4. Responsibilities

As per the roles they have traditionally occupied within the research system, researchers, research communities, research institutions and research funders share the responsibilities and costs of ensuring a robust and open research data environment in Canada. Therefore, they should work collaboratively in addressing the gaps in human and technical infrastructure to meet this objective. This section outlines their responsibilities in meeting the expectations described in Section 3.

Responsibilities of **researchers** include:

- incorporating data management best practices into their research;
- developing data management plans to guide the responsible collection, formatting, preservation and sharing of their data throughout the entire lifecycle of a research project and beyond;
- following the requirements of applicable institutional and/or funding agency policies and professional or disciplinary standards;
- acknowledging and citing datasets that contribute to their research; and
- staying abreast of standards and expectations of their disciplinary community.

Responsibilities of **research communities** include:

- developing data management standards, promoting and communicating existing standards to ensure that they are used, and working collaboratively to review and improve these standards;
- recognizing data as an important research output and fostering excellence in data management within their research community; and
- identifying, promoting and encouraging the use of repositories and platforms that meet or exceed data management standards.

Responsibilities of **research institutions** include:

- providing their researchers with an environment that enables world class data stewardship practices;
- delivering, or supporting access to, repositories or other platforms that securely preserve, curate and provide continued access to research data;
- supporting researchers in their efforts to establish and implement data management practices that are consistent with ethical, legal and commercial obligations, as well as tri-agency requirements, including the [Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans](#) – 2nd edition, the [Tri-Agency Framework: Responsible Conduct of Research](#) and other relevant policies;
- providing their affiliated researchers with guidance to properly manage their data in accordance with both the principles outlined above and research community best practices, including the development of data management plans;
- recognizing data as an important research output and fostering excellence in data management;
- promoting the importance of data management to researchers, staff and students; and
- developing their own data management policies, ensuring that these policies are in accordance with the principles outlined above and provincial and national laws, and can accommodate the rapidly evolving research communities' best practices.

Responsibilities of **research funders** include:

- developing policies and requirements that enable and recognize responsible data management, in accordance with the principles outlined above;
- providing applicants with clear information and guidance with regard to fulfilling data management requirements;
- recognizing data as an important research output;
- promoting the importance of excellent data management; and
- where appropriate, providing peer reviewers with guidance and developing assessment material for including data management considerations in the application assessment process.