

RESEARCH DATA MANAGEMENT IN CANADIAN UNIVERSITIES

A STATEMENT OF PRINCIPLES

MARCH 2016

Background

This statement of research data management (RDM) principles has been developed by a task group of senior university administrators and individuals from Research Data Canada (RDC) as follow-up to a meeting in Edmonton on November 17, 2015 co-sponsored by RDC and the University of Alberta with support from SSHRC on behalf of the federal granting agencies. The Edmonton meeting involved senior representatives of 16 institutions, as well as of the federal granting councils and other national organizations with an interest in research data. All who gathered in Edmonton were cognizant of RDM issues in part because of work underway to develop a set of guidelines on data management by the federal granting agencies. <http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1>

Those gathered in Edmonton delegated the development of a statement of principles to a task group¹. The group started its deliberations with a review of several comparable statements from around the world before selecting as a starting point the principles developed by the Research Councils of the United Kingdom <www.rcuk.ac.uk/research/datapolicy>. The authors of the current document acknowledge their debt to the RCUK.

Universities will develop RDM policies to suit institutional needs and circumstances, but they can learn from one another in a collaborative environment that avoids each one “reinventing the wheel”. If they all share the core principles articulated below, joint development of new tools, services, and infrastructure will be significantly enhanced and the objectives of open science and collaborative data sharing will be advanced. Some universities have already developed institutional policies dealing with RDM. Others are in the midst of such policy development, and still others have yet to begin the process. The purpose of this statement of principles is not to create one national RDM policy, nor to

¹ The task group included the following: Mark Leggott, University Librarian, University of Prince Edward Island (Research Data Canada Executive Director Designate); Frederic Bouchard, Vice-recteur associé, Université de Montréal; Robert Haché, Vice-President Research and Innovation, York University; Gerald Beasley, Vice-Provost and Chief Librarian, University of Alberta; Susan Babcock Acting Executive Director Research Ethics Office, University of Alberta; David Castle, Vice-President Research, University of Victoria; Andrew Bjerring, Chair RDC Policy Committee; Walter Stewart, RDC Co-ordinator.

replace existing institutional policies, but rather to try to ensure that there is a common core to the principles governing RDM among Canada's universities.

CANADIAN UNIVERSITY RESEARCH DATA MANAGEMENT STATEMENT OF PRINCIPLES

Preamble

As part of their dedication to research excellence and knowledge dissemination, Canadian universities are committed to meeting the challenges and opportunities involved in the effective management of research data. These principles are stated in a manner consistent with the glossary of terms developed by Research Data Canada, <http://www.rdc-drc.ca/glossary>.

Specifically, research data means:

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 and/or are commonly accepted in the research community as necessary to validate research findings and results. All other digital and non-digital content have the potential of becoming research data. Research data may be experimental data, observational data, operational data, third party data, public sector data, monitoring data, processed data, or repurposed data.

Metadata means:

Literally, “data about data”; data that defines and describes the characteristics of other data, used to improve both business and technical understanding of data and data-related processes. Business metadata includes the names and business definitions of subject areas, entities and attributes, attribute data types and other attribute properties, range descriptions, valid domain values and their definitions. Technical metadata includes physical database table and column names, column properties, and the properties of other database objects, including how data is stored. Process metadata is data that defines and describes the characteristics of other system elements (processes, business rules, programs, jobs, tools, etc.). Data stewardship metadata is data about data stewards, stewardship processes and responsibility assignments.

Research Data Management (RDM) means:

RDM – Data Management refers to the storage, access and preservation of data produced from a given investigation. Data management practices cover the entire lifecycle of the data, from planning the investigation to conducting it, and from backing up data as it is created and used to long term preservation of data deliverables after the research investigation has concluded. Specific activities and issues that fall within the category of Data Management include: File naming (the proper way to name computer files); data quality control and quality assurance; data access; data documentation (including levels of uncertainty); metadata creation and controlled vocabularies; data storage; data archiving & preservation; data sharing and re-use; data integrity; data security; data privacy; data rights; notebook protocols (lab or field).

The Principles

1. The Importance of Data for Research:

Canadian universities recognize the central role of data in 21st century research. Data are both the product of research and a foundation for future research. The appropriate management of research data and facilitating appropriate access to that data are foundations of modern scholarship and scientific discovery.

2. National and International Collaboration:

Canadian universities will aim to meet RDM challenges and opportunities collaboratively and in alignment with international activities in RDM, including the development and setting of standards such that Canadian research data are interoperable with those of global research partners.

3. Access:

Research data generated by publicly-supported research are a public good, produced in the public interest, and will normally be made openly available in a timely and responsible manner with as few restrictions as possible. Rights to publicly supported research data will not normally be assigned to others without ensuring that the data remain accessible and available for reuse.

4. Ethical, Legal, and Privacy Issues:

Canadian universities recognize that there are privacy considerations, legal concerns, ethical issues, and commercial interests reflected in contractual requirements that may constrain the release of research data. Institutional policies will be developed to accommodate these complexities.

5. Privileged Use:

Those who conduct research are entitled to a limited period of privileged use of the data they have collected and generated, for example to enable them to publish the results of their research. Such limited periods may vary in length depending upon the academic discipline involved.

6. Recognition of Intellectual Contributions:

Institutional policies and reward systems will recognize the intellectual contributions of researchers who generate, preserve, and share research data. Users of already generated research data are obliged to acknowledge the source of their data and abide by the terms and conditions under which they are accessed.

7. A Public Trust:

As research data constitute a public good, appropriate management of such data constitutes a public trust. It is recognized that while RDM may be the primary responsibility of the principal investigator during the life of the research project, long term preservation is only possible with appropriate funding from funding agencies and governments. Resources will be commensurate with the original support provided to the research, including training and expertise.

8. Data Management Plans:

Institutional and project-specific data management plans typically follow recognized, relevant international standards and community best practices. Such plans should recognize that data may be of potential long-term value, sometimes for purposes distinct from those for which the data were created, and will require plans and resources for preservation and access. Decisions about the length of time for data preservation should be based on policies which recognize the potential long-term value of research data.

9. Metadata and Discoverability:

Metadata will normally be recorded and made openly available in an internationally recognized standard. This will enable research data to be discoverable and effectively re-used by others. Published results will include information on how to access the data on which the results are based. If the data cannot be or is not yet available (see principles 4 and 5), the metadata may be published in order to alert potential users to the existence of such data.

10. Multilingual Access:

Canadian scholars and institutions may employ tools for data management in either French or English, ideally both.