Background Reading:

Implementing Research Data Management Policies

Prepared by Kathleen Shearer

On behalf of the Research Data Canada Policy Subcommittee

May 2014

Contents

About Research Data Canada

Research Data Canada is a collaborative effort to address the challenges and issues surrounding the access and preservation of data arising from Canadian research. This multi-disciplinary group of universities, institutes, libraries, granting agencies, and individual researchers has a shared recognition of the pressing need to deal with Canadian data management issues from a national perspective.

1. National Approaches to Policy Implementation

Funding agencies, universities and research institutions around the world are implementing policies that govern how research data produced are managed and shared. According to the introduction to the University of Edinburgh's "Policy for Research Data Management":

> "At international, national and local levels, there is intense interest in how to manage the rapidly expanding volume and complexity of research data. Concern is both for the shorter term –ensuring competitive advantage through secure and easy-to-use access, and for the longer term – ensuring enduring access and usability to the research community into the future and compliance with legislation."[1]

No single player in the research ecosystem system can ensure the stewardship of research data. For data to be accessible over the long-term, a continuous line of responsibility for maintaining data throughout their lifecycle is required.

A review of existing policies by the Research Data Canada Policy Committee found that there are different approaches to policy language and implementation. This is in large part there are varying levels of infrastructure support available and a reticence amongst researchers amongst some communities. Below are brief descriptions of the national contexts for policy development in three jurisdictions: Australia, United Kingdom and United States.

Australia

In Australia, the Australian Code for Responsible Conduct of Research[2] places the onus of responsibility on the universities. Requiring institutions to retain research data, provide secure data storage, identify ownership, and ensure security and confidentiality of research data. Researchers are required to retain research data and primary materials, manage storage of research data and primary materials, maintain confidentiality of research data and primary materials.

The Australian government, through Australian National Data Service (ANDS), has provided funding and support for Australian Universities in order to support the implementation of the Code. Several universities in Australia have adopted data management policies, based on the Code. However, at the time of review in Fall 2013, institutional policies were rather vague in terms of the institutional role in managing research data. For example, Monash University's Research Data Management Policy states that "research data and materials must be stored securely to protect against theft, misuse, damage or loss. Research data must be held in appropriate facilities that allow access to be managed as required."[3] However, there is no reference to what might constitute an 'appropriate facility'.

[1] http://www.docs.is.ed.ac.uk/docs/data-library/rdm-policy.pdf
[2] http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39.pdf
[3] http://www.policy.monash.edu/policy-bank/academic/research/research-data-management-policy.html

United Kingdom

The UK likely has the most comprehensive policy approach to data management of all the jurisdictions reviewed. Each of the eight federal Research Councils has a data sharing and management policy. Policy requirements for each council are different and, while the funding councils maintain a number of disciplinary data centres, these do not have the scope and capacity to collect and make available all research data produced through council funding. The Digital Curation Centre (DCC) says, "For research that falls outside subject data centre remits, the institutions in which funded researchers are based are expected to maintain outputs in the long-term."[4]

United States

In the US, both the National Science Foundation (NSF) and the National Institutes of Health (NIH) have data sharing policies. The policies are not prescriptive in terms of responsibilities for long-term management of the data but rather require that research teams develop data management plans that describe how the data will be preserved and shared.

Since 2011, the NSF has required that all proposals include a supplementary document of no more than two pages labeled Data Management Plan (DMP). The plan should describe how the research teams will conform to the policy. The NIH Policy, which also requires a plan for data management and sharing only applies to research projects that have budgets of $500,000 US per year or more.

[4] http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

2. Comparison of research data management policies

Table 1: Comparison of research funders research data management policies

	NIH	NSF	ESRC (UK)	Wellcome Trust
Coverage	All applicants seeking $500,000 or more in direct costs in any year of the project period.	All investigators	All investigators	All investigators
Time Limits	No later than the acceptance for publication of the main findings from the final dataset	"within a reasonable time	Data must be made available for preparation for re-use and/or archiving within three months of the end of the award.	Published outputs should be deposited as soon as possible, and in any event within six months of final publication.
Data management plan	Investigators "are expected to include a plan for data sharing or state why data sharing is not possible"	Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan".	Grant applicants are required to submit a statement on data sharing in the relevant section of the Je-S form and provide a c.2 page data management and sharing plan.	Researchers are required to submit a plan for data management and sharing, where a proposal involves the generation of datasets that have clear scope for wider research use and hold significant long-term value.
Dissemination/ sharing	The NIH expects and supports the timely release and sharing of final research data from NIH-supported studies for use by other researchers.	Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants.	Research data should be made available to the scientific community in a timely and responsible manner. The data service supports data sharing.	Researchers are to maximise the availability of data with as few restrictions as possible.
Preservation	-----Nothing-------	------Nothing-------	The ESRC data service providers are responsible for ensuring long-term access to the data.	Institutions are expected to have guidelines setting out responsibilities and procedures for the appropriate storage and disposal of data and samples. Data should be

				maintained securely for a minimum of 10 years.
Monitoring	--------Nothing------	-------Nothing------	The final payment of a grant may be withheld if data has not been offered for deposit to the required standard, unless a waiver has been agreed in advance.	All awardees are asked to report back on their approach for disseminating their research as part of their end of grant report.

Institutional Example: University of Edinburgh Research Data Management Policy[5]

This policy for managing research data was approved by the University Court on 16 May 2011.

The University adopts the following policy on Research Data Management. It is acknowledged that this is an aspirational policy, and that implementation will take some years.

1. Research data will be managed to the highest standards throughout the research data lifecycle as part of the University's commitment to research excellence.
2. Responsibility for research data management through a sound research data management plan during any research project or programme lies primarily with Principal Investigators (PIs).
3. All new research proposals [from date of adoption] must include research data management plans or protocols that explicitly address data capture, management, integrity, confidentiality, retention, sharing and publication.
4. The University will provide training, support, advice and where appropriate guidelines and templates for the research data management and research data management plans.
5. The University will provide mechanisms and services for storage, backup, registration, deposit and retention of research data assets in support of current and future access, during and after completion of research projects.
6. Any data retained elsewhere, for example in an international data service or domain repository should be registered with the University.
7. Research data management plans must ensure that research data are available for access and re-use where appropriate and under appropriate safeguards.
8. The legitimate interests of the subjects of research data must be protected.

[5] From: http://www.ed.ac.uk/schools-departments/information-services/about/policies-and-regulations/research-data-policy

9. Research data of future historical interest, and all research data that represent records of the University, including data that substantiate research findings, will be offered and assessed for deposit and retention in an appropriate national or international data service or domain repository, or a University repository.
10. Exclusive rights to reuse or publish research data should not be handed over to commercial publishers or agents without retaining the rights to make the data openly available for re-use, unless this is a condition of funding.

3. Challenges for policy implementation

In the report, *Riding the Wave*, the European Commissions' High-Level Expert Group on Scientific Data outlines some of the issues involved in achieving widespread sharing and preservation of research data:

> "But there are many challenges. How can we organise such a fiendishly complicated global effort, without hindering its flexibility and openness? How do we incentivise researchers, companies, and individuals to contribute their own data to the e-infrastructure – while still trusting that they can protect their privacy or ownership? How can we manage to preserve all this data, despite changing technologies and needs? How to convey the context and provenance of the data? How to pay for it all?"[6]

Indeed there are a number of well-acknowledged challenges for institutions and funding agencies looking to implement research data management policies that are outlined in more detail below. These challenges are by no means insurmountable, but organizations will want to consider how to address these challenges in advance.

Researchers' attitudes

Researchers' unwillingness to share is one of the major obstacles to implementing data management policies. Researchers often have a strong sense of ownership of their data and surveys have found that they are concerned about being scooped if they make their data public or that they will not be given due credit.[7] Time is another big issues for researchers. Data sharing requires that researchers prepare their data for others to understand and re-use, often a very time consuming process, especially if this has to be done at the end of the research project. Data management plans can address this issue, as they help to ensure that researchers assign appropriate metadata and standards to the data collection/production phase.

Infrastructure and skills

For research data to be available after the lifespan of a specific research project, they must be integrated into an enduring institutional environment supported by a digital repository. In Canada (as elsewhere) there are still numerous gaps in this infrastructure. There are some large international discipline-based repositories in certain fields as well as data repositories that are maintained by national agencies, however they cover only a small portion of the research data produced in Canada. There are a number of stakeholders looking now at how to fill these gaps. A federated pilot project lead by Research Data Canada is looking at how to expand existing infrastructure to serve a greater number of researchers. In addition, a library-based research data management network is being developed by CARL and the four regional academic library associations.

---

[6] *Riding the Wave. Final report of the High-Level Expert Group on Scientific Data*. (2010) European Commission. October 2010. Available at: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf

[7] Broadley, Philippa and Kerry Raymond. (2012) *Researcher Attitudes to Data Sharing: Cultural Change Requires Better Motivations.* Available at: http://eresearchau.files.wordpress.com/2012/08/eresau2012_submission_42.pdf

Skills and knowledge of data management are another important requirement for data sharing. Data stewardship requires the active management of data over its lifecycle and involves activities such as "appraising, selecting, depositing or ingesting data into a repository, ensuring authenticity, managing the collection of data and metadata, refreshing digital media, and migrating data to new digital media."[8] In order to comply with research data policies, therefore, researchers will need access to services that provide support for RDM. These services do exist at some institutions and in some disciplines, but not comprehensively across the country. In order to ensure there is better support for research data management across Canada, RDC and the library based Project ARC initiative are developing a plan for the implementation of a centre of expertise for RDM in Canada which would provide training, resources and consultation services for both institutions, researchers and the library community.

Complex policy environment

There are a number of related policies at the local and national level that overlap with some of the elements that might be included in a research data management policies. Table 2 provides an overview of these policies and Appendix 2 lists the sections in these policies that are relevant to research data management. This can sometimes create a confusing environment for those wishing to adhere to RDM policies. For example, researchers working with data related to human participants or other types of confidential data are often unsure whether they can adhere to both policies concurrently. A narrow interpretation of Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans (TCPS) by Research Ethics Boards (REBs) or researchers can result in the unnecessary destruction of data related to human subjects in contravention with data sharing policies.

[8] Research Data Strategy Working Group. (2008) *Stewardship of Research Data in Canada: A Gap Analysis*. Available at: *rds-sdr.cisti-icist.nrc-cnrc.gc.ca/docs/GapAnalysis.pdf>*

4. Intersecting policies

Table 2: Policies related to research data management

	Federal & provincial governments	Funding Agencies	Universities and research institutions	Research projects	Data archives, centres and repositories	Computing facilities
Privacy Act (1)	X					
Personal Information Protection and Electronic Documents Act (2)	X					
Copyright Law	X					
Open Data Policy (no policy yet in Canada)	X					
TCPS 2—2nd edition of Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (3)		X				
Code of Conduct/ Research Integrity Policy (4)			X			
Data Management Policy (5)		X	X	X		
Data Collection Policy (6)					X	
Data Storage Policy (7)						X

5. Generic policy elements

Data Management Plans: Research proposals must include a Data Management Plan in proposal.

Data Quality and Standards: Investigators are required to adhere to international standards to enable access and reuse in the discipline.

Data Access: Data documentation and metadata must accompany data so that the data is understandable by others.

Data Access/Visibility: Investigators are required to either (1) deposit data in relevant subject or institutional repositories which promotes data visibility and facilitates access to authenticated users; or, (2) where there is no repositories hold the data locally, and make it available through a web-based presence (e.g. study or laboratory website) which promotes data visibility and facilitates access. (This may include a high-level study description and inventory of key datasets that informs potential.

Data Access: Investigators must make data accessible by one of above methods in a timely manor-usually upon acceptance of publication.

Privacy: In the case of data about human subjects, investigators are required to adhere to those policies (see below). The rights and privacy of individuals who participate in research must be protected at all times. Thus, data intended for broader use should be free of identifiers that would permit linkages to individual research participants and variables that could lead to deductive disclosure of the identity of individual subjects.

Data Ownership: Issues of data ownership can arise when co-funding is provided by the private sector (e.g., the pharmaceutical or biotechnology industries) with corresponding constraints on public disclosure. The organization recognizes the need to protect patentable and other proprietary data. Any restrictions on data sharing due to co-funding arrangements should be discussed in the data-sharing/management plan section of an application and will be considered by program staff. The organization recognizes an institution's desire to exercise its intellectual property rights may justify a need to delay disclosure of research findings, a delay of 30 to 60 days is generally viewed as a reasonable period for such activity. (NIH)

Data Retention: Investigator must have clearly defined responsibility for recording, retaining, and storing research data. These records should include sufficient detail to permit examination for the purpose of replicating the research, responding to questions that may result from unintentional error or misinterpretation, establishing authenticity of the records, and confirming the validity of the conclusions. Data should be retained for a certain time limit (on average 5 year)

Data Preservation: Investigators must deposit their data in a long-term archive to ensure the preservation of their data.

Exceptions to policies

- The rights and privacy of individuals who participate in research must be protected at all times.
- Where local and traditional knowledge is concerned, rights of the knowledge holders shall not be compromised.
- Where data release may cause harm, specific aspects of the data may need to be kept protected (for example, locations of nests of endangered birds or locations

of sacred sites).

- The organization recognizes that it may be necessary on occasion to delay publication for a short period to allow time for applications to be drafted.

Appendix 1: Selective guidance for policy implementation

A number of recommendations for organizations intending to implement policies were identified based on this review, documented below:

- Understand the cost implications of adoption of policies
- Agree on how costs and responsibilities will be distributed across stakeholders
- Develop methods for monitoring adherence
- Implement incentives for researchers to use appropriate standards and deposit their data
- Ensure there is consistency and harmonization of policies across regions and organizations
- Decide on the nature of the repository infrastructure (i.e. centralized vs. community-based vs institutional- or a mixture)
- Researcher engagement and raise awareness of the policy with the research community and other stakeholders
- Provide assistance for researchers to understand and adhere to policy
- Develop a policy based on high-level principles
- Start with achievable objectives: i.e. requirements that researchers attach a data management plan to funding applications
- Develop an implementation roadmap and secure necessary resources.
- Be aware of the scale and nature of the data to be managed.
- Share as much information as you can so that the community is well informed and can engage with the process.
- Phase in the policy, and avoid strict deadlines

Appendix 2: Extracts from policies related to research data management

1. **Privacy Act:** A government institution shall dispose of personal information under the control of the institution in accordance with the regulations and in accordance with any directives or guidelines issued by the designated minister in relation to the disposal of that information.[9]

2. **Personal Information Protection and Electronic Documents Act:** Personal information shall not be used or disclosed for purposes other than those for which it was collected, except with the consent of the individual or as required by law. Personal information shall be retained only as long as necessary for the fulfillment of those purposes.

   Personal information that is no longer required to fulfill the identified purposes should be destroyed, erased, or made anonymous. Organizations shall develop guidelines and implement procedures to govern the destruction of personal information.[10]

3. ***Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans:*** Researchers should ensure that the data obtained are stored with all the precautions appropriate to the sensitivity of the data. Information that identifies individuals or groups should be kept in different databases with unique identifiers.

   Researchers' plans for preserving or destroying participants' data must be appropriate to the field of research and the wishes of participants. For example, in oral history the best practice may be to archive the information collected (with the participants' consent) for future generations. With research where the release of information could harm participants, it may be best to destroy the data collected as soon as possible.

   Explain your plans for preserving and protecting participants' data or for destroying data in light of the best practices in your field of research and the wishes of participants. Some funding agencies, professional organizations and publishers have established minimum requirements for data retention (e.g., five years), after which time the data are to be destroyed. You must disclose their plans for data destruction that includes a time frame and the methods that will be employed to destroy the data (e.g., shredding, electronic file deletion).[11]

4. **Guidelines for Ethical Practices in Research:** Generally accepted scholarly standards and practices include: 2.1.4 keeping complete and accurate records of data, methodologies and findings, including graphs and images, in a manner that will allow verification or replication of the work by others. This includes recording all primary data in clear, adequate, original and chronological form, and retaining it in a repository from which it cannot be removed. The principal investigator is responsible for the collection, maintenance and retention of research data. Records of data should normally be retained in the unit in which they are produced for at least five years after the work is published or otherwise presented (if the form of the record or data permits this, and subject to any assurances that data would be destroyed to assure anonymity). Records should

[9] http://laws-lois.justice.gc.ca/eng/acts/P-21/page-2.html
[10] http://laws-lois.justice.gc.ca/eng/acts/p-8.6/page-19.html
[11] http://www.pre.ethics.gc.ca/eng/archives/tcps-eptc/section3-chapitre3/

be retained in their original medium, or transferred to a secondary medium provided that the transfer process is fully validated, and the person who transfers the data from the original to the secondary medium attests that the secondary documents are true copies of the respective original record including any and all notations, corrections, or other changes made to the original record prior to the creation of the secondary documents. In the case of collaborative research all those involved in the conduct of the research will have access to the data (subject to any assurances that access to the data would be restricted to assure anonymity), and will normally be allowed to make copies of the record.[12]

5. **Data Management Policy** for funding agencies, institutions, and research projects

These policies may contain the following elements:

5.1 Data Management Plans: Research proposals must include a Data Management Plan.

5.2 Data Quality and Standards: Investigators are required to adhere to international standards to enable access and reuse in the discipline.

5.3 Data Access: Data documentation and metadata must accompany data so that the data is understandable by others.

5.4 Data Access/Visibility: Investigators are required to either (1) deposit data in relevant subject or institutional repositories which promotes data visibility and facilitates access to authenticated users; or, (2) where there is no repositories hold the data locally, and make it available through a web-based presence (e.g. study or laboratory website) which promotes data visibility and facilitates access. (This may include a high-level study description and inventory of key datasets that informs potential.

5.5 Data Access: Investigators must make data accessible by one of above methods in a timely manor-usually upon acceptance of publication.

5.6 Data Ownership: Issues of data ownership can arise when co-funding is provided by the private sector (e.g., the pharmaceutical or biotechnology industries) with corresponding constraints on public disclosure. The organization recognizes the need to protect patentable and other proprietary data. Any restrictions on data sharing due to co-funding arrangements should be discussed in the data-sharing/management plan section of an application and will be considered by program staff. The organization recognizes an institution's desire to exercise its intellectual property rights may justify a need to delay disclosure of research findings, a delay of 30 to 60 days is generally viewed as a reasonable period for such activity. (NIH)

5.7 Data Retention: Investigator must have clearly defined responsibility for recording, retaining, and storing research data. These records should include sufficient detail to permit examination for the purpose of replicating the research, responding to questions that may result from unintentional error or misinterpretation, establishing authenticity of the records, and confirming the validity of the conclusions. Data should be

[12] http://universitycounsel.ubc.ca/files/2013/04/policy85.pdf

retained for a certain time limit (e.g. 5 year)

5.8 Data Preservation: Investigators must deposit their data in a long-term archive to ensure the preservation of their data.

5.9 Exceptions to data management policies

5.9.1 The rights and privacy of individuals who participate in research must be protected at all times.

5.9.2 Where local and traditional knowledge is concerned, rights of the knowledge holders shall not be compromised.

5.9.3 Where data release may cause harm, specific aspects of the data may need to be kept protected (for example, locations of nests of endangered birds or locations of sacred sites).

5.9.4 The organization recognizes that it may be necessary on occasion to delay publication for a short period to allow time for applications to be drafted.

6. Repository collection policy (data repository)

May include[13]:

6.1 **Submission policies**: Who can deposit? What type of materials can they deposit and in what format? What level of moderation is required for checking deposits, if any?

6.2 **Collection policies**: will the repository focus on a specific discipline, or will it reflect the entire academic output? What types of materials are sought? What metadata must be collected? What versions are acceptable? Should peer or quality reviews be implemented?

6.3 **Preservation policies**: for how long will the repository aim to preserve deposits? Can this be guaranteed? What formats should be used for preservation purposes?

6.4 **Usage policies**: what can end-users and services do with repository metadata and content? How should publishers' restrictions or embargoes be managed? At what level should usage be permitted, e.g. on an item-by-item level? Is there a takedown policy to respond to copyright or other infringements?

7. Data Storage Policy: All data will be deleted from a user's directories once their WestGrid account has been retired (this occurs 90 days after the account has been deactivated). See the User Accounts pages for more information about the deactivation process. Some data may remain on backup tapes for a period of time. Any recovery of the user's data after account retirement is strictly on a best-effort basis with no guarantee of success.[14]

[13] http://www.rsp.ac.uk/documents/briefing-papers/repoadmin-policyv2.pdf
[14] https://www.westgrid.ca/resources_services/data_storage/data_storage_policy