

## **CARL Response to the Draft Tri-Agency Statement of Principles on Digital Data Management**

*September 3<sup>rd</sup>, 2015*



**Canadian Association  
of Research Libraries**

**Association des bibliothèques  
de recherche du Canada**

## ***Who we are***

The Canadian Association of Research Libraries (CARL) represents the country's twenty-nine largest university libraries. Enhancing research and higher education are at the heart of our mission. CARL promotes effective and sustainable scholarly communication, and public policy that enables broad access to scholarly information.

## ***CARL Response to the Draft Tri-Agency Statement of Principles on Digital Data Management***

The Canadian Association of Research Libraries (CARL) is delighted to respond to the Tri-Agency Statement of Principles on Digital Data Management.

CARL's Portage Network supports a national research data management (RDM) service to assist researchers and other RDM stakeholders through a library-based network of expertise on RDM and through national platforms for planning, preserving, and discovering research data.

We strongly support the Tri-Agency Statement of Principle for Digital Data Management which will provide a guiding framework for the development of more concrete policies and practices in Canada. We are pleased that both expectations and responsibilities are articulated in the Statement of Principles. The ecosystem metaphor for data stewardship is popular because it expresses the interdependence of the many stakeholders engaged in research data infrastructure. How these stakeholders behave in this ecosystem, however, remains rather open-ended at this time and we feel that both the expectations and responsibilities will help define the collaborative roles required of RDM stakeholders in Canada.

## **Expectations**

We have some comments about each of the expectations discussed in the Statement and hope that you see this as providing you with a perspective from an organization committed to research data management services and infrastructure.

- i. **Data management planning.** Such plans need to occur over several levels of organization, covering data activities at the project or researcher level, at the institutional level, and at a regional or national consortial level. Consequently, the information in a data management plan needs to be collectively shared among all of those engaged in the stewardship of data. Data management plans can serve as a tool of collaboration among stakeholders in the research data ecosystem.

While it is vital to get researchers to identify the costs, benefits, and challenges of managing data, they should be able to do this in the context of institutional and national research data services and infrastructure. Portage is an example of one national service that offers assistance to researchers, including its web-based tool for data management plans.

We recommend adding the following sentence to this section of the Statement: “**Research plans should be developed using standardized tools where available, and shared with other parties, including funders and those responsible for long-term data stewardship.**”

- ii. **Constraints and obligations.** Interpreting how legal and ethical constraints and obligations impinge upon data management practices is an area of confusion for researchers as well as those providing data support to researchers. A primary problem with legal and ethical obligations is around how they get interpreted when applied to data management. Researchers should not require a lawyer to help them understand the legal and ethical treatment of research data. Portage is committed to addressing this problem by developing specialized knowledge about legal and ethical obligations regarding data management within its network of expertise.
- iii. **Adherence to standards.** Finding guidance on the appropriateness of data management standards can be a formidable task for a researcher. While the best practices identified within a domain should be initially consulted, gaps exist within and between disciplines about what standards to employ. The Portage network of expertise will be preparing information resources to direct researchers toward relevant solutions.
- iv. **Collection and storage.** Another challenge for researchers is the selection of software and formats best suited for analysis, storage, sharing, and preservation. In many instances, separate formats are required for these different functions. Portage is developing a Format Policy Registry to help researchers identify appropriate data formats and to support machine-actions that require information about formats, such as, transforming a format for data sharing or preservation.
- v. **Metadata.** The challenge for the researcher regarding metadata is once again about having enough relevant information to make a choice and about having access to tools that support metadata production across the data lifecycle. Part of the Portage network of expertise will provide specialized consultation on metadata, describing the choices that are available and the tools to assist in its production.

Furthermore, metadata are instrumental in the discovery and re-use of research data. We feel that this should be stated explicitly in the principles: “**Quality metadata are essential for the discovery of research data and for the systems that exploit or mine the data.**”

- vi. **Preservation, retention, and sharing.** Preservation is critical for ensuring that research data are available for sharing and reuse in the future. Preservation is a complex endeavour involving planning, resource allocation, and application of preservation practices in the context of an enduring institutional environment. While the number of data repositories in Canada and elsewhere in the world is increasing, the current count of *publicly curated repositories* is insufficient to meet the demands for preserving research

data. Therefore, for now, this principle is largely aspirational. That being said, those providing research data management infrastructure are assembling and testing platforms that will increase the capacity for sharing and preserving data. Portage will partner with other infrastructure providers to support publicly curated repositories and will provide information about them to assist researchers in making an appropriate choice for their data.

Policies and practices around data retention require community-approved criteria upon which to evaluate long-term retention. Without such criteria, looking at cost effectiveness could too easily become the only factor in determining which data are preserved.

The current set of principles fail to address data management practices required for sensitive data, especially data about human subjects. We suggest that this can be resolved by including the following statement: “***The sensitive nature of some research data prevent it from being shared in its entirety. Techniques do exist, however, for de-identifying or making this type of data safe to share more widely with others. In some instances, sensitive data are made available through enclave facilities that control who has access to the data and for what purposes.***”

- vii. **Timeliness.** We fully support the principle of linking data to relevant publications and see the need for technologies enabling interoperability between the discovery of scholarly communications and access to the data upon which these communications are based. Portage will work with the library community to implement the technologies needed to support this interoperability.

We agree that researchers should share their data as soon as possible in a manner that respects any privacy or security constraints and no later than upon the publication of results. In terms of a defined period for exclusive use of data for primary research, commonly referred to as an embargo period, this should not be longer than the time needed for the original data producers to publish their results.

- viii. **Acknowledgement and citation.** This is a topic on which the library community has been a long-standing advocate and provider of solutions. We strongly support the position in the Statement saying, “Researchers who responsibly and effectively share their data should be recognized by funders, their academic institutions and users benefiting from the re-use of the data.” We reinforce this point below under our suggested improvements to the responsibilities of institutions and funders. Furthermore, the library community can assist in the gathering of the information required to calculate the metrics used to assess data use through citation.
- ix. **Efficient and cost effective.** Clearly, costs and benefits are an important component in planning project-level data management. However, significant cost savings are more likely to be found at the data management infrastructure level. The Statement’s principle of efficiency and cost effectiveness should also emphasize the need for collaboration among

data management infrastructure providers within the data ecosystem, working together to achieve efficient operations that are cost effective while maintaining excellence in service.

## **Responsibilities**

Key to the proper stewardship of research data is ensuring that roles and responsibilities are assigned and acknowledged across the various stakeholder communities. We would like to make two points about the list of responsibilities and the four stakeholder groups who are identified in the Statement.

First, several of these responsibilities are shared among stakeholders. We feel that shared responsibilities should be listed separately as collective responsibilities and that more specific aspects of these shared responsibilities should be identified for individual stakeholders. For example, recognition for the importance of data as a research output could be re-expressed for researchers, institutions, and funders as follows:

- i. for researchers, through providing proper attribution to the data that they use from other researchers;
- ii. for research institutions, through developing rewards for data as a research output, such as incorporating data activity metrics in their tenure and review process, and
- iii. for research funders, through incentives in future grant submissions from researchers who have deposited research data with a recognized repository.

In addition, “recognizing the importance of data as a research output” and “promoting the importance of data management” are also shared responsibilities across all communities. These are examples of responsibilities that could be assumed by different stakeholders sharing a common overall responsibility.

Second, some stakeholders bridge the categories identified in the Statement or provide unique research data management services that should be added to the current list of stakeholders. Portage is an example in that it represents multiple research institutions and is part of the research community. But we also see Portage as having responsibilities that warrant further categorization. There are other significant stakeholders in the same situation, such as CANARIE, Compute Canada, and Research Data Canada. Together these organizations represent the ever-evolving Canadian research data infrastructure or what Industry Canada is calling the Digital Research Infrastructure. The responsibilities of the organizations in the research data infrastructure group include:

- i. working together and with the other stakeholders to provide a comprehensive set of data management services and infrastructure;
- ii. coordinating the delivery of research data infrastructure with regional and national services;

- iii. supporting national data stewardship through a network of data repositories; and
- iv. ensuring a national level of access to research data infrastructure for researchers irrespective of institutional size or location.

***We feel that the next draft of this Statement should include a category of responsibilities for Research Data Infrastructure Providers.***

In terms of specific comments on the responsibilities, we suggest that you make a few small revisions to improve the language:

- For researchers: “Following the requirements of applicable institutional and/or **funding agency policies**, and professional or disciplinary standards.
- For research communities: “Identifying and encouraging the use of repositories and platforms **that meet or exceed established data management standards.**”

CARL and Portage are committed to the Tri-Agency’s principles on digital data management. We are grateful to the Tri-Agency for drafting these expectations and responsibilities and for seeking community feedback. A periodic review conducted in conjunction with the stakeholders in the research data ecosystem would be welcomed. Obtaining stakeholder commitment to these principles will be important in order to firmly establish data stewardship in Canada.

Furthermore, we look forward to these principles being used to underpin data management policy requirements in the near future from each of the agencies, and hope that, like other policies, these requirements will be harmonized across the agencies over time.

## Canadian Association of Research Libraries

### **CARL President**

Martha Whitehead,  
Vice Provost and University Librarian,  
Queen's University

### **CARL Executive Director**

Susan Haigh  
[susan.haigh@carl-abrc.ca](mailto:susan.haigh@carl-abrc.ca)

Tel. 613.482.9344    203-309 Cooper St.  
[info@carl-abrc.ca](mailto:info@carl-abrc.ca)    Ottawa, ON  
[www.carl-abrc.ca](http://www.carl-abrc.ca)    K2P 0G5

