

CARL Response to Canada's Fundamental Science Review: Call for Evidence and Input

SEPTEMBER 30, 2016

203-309 Cooper St.
Ottawa ON K2P 0G5
613.483.9344

www.carl-abrc.ca



Table of Contents

Introduction.....	2
Program Coordination	2
Balance Across Funding Elements.....	3
Supporting the Needs of Multidisciplinary Research	5
International Programs	6
Other Comments	6

Introduction

CARL is please to submit comments and recommendations to Canada's Fundamental Science Review: Call for Evidence and Input¹ by focusing on the following questions;

1. Coordination between the programs being provided by the granting councils and other funding organizations, provinces, and/or amongst themselves. Are there areas for improvement?
2. Is there a need for the federal government to improve the balance across funding elements (e.g. investments in principal researchers, funding of research staff and other direct costs of research, funding of infrastructure and equipment operations and maintenance, and reimbursement of indirect costs)? If so, how can this balance be achieved? What is the appropriate federal role in supporting infrastructure operating costs? Do CFI and granting councils programs work in a complementary fashion?
3. Are current federal programs supporting the needs of multidisciplinary research programs? If not, how can the situation be improved? Does the funding ecosystem (funding councils and other agencies) work collaboratively and effectively across disciplines?
4. Are there any other issues or questions that you would like to raise and address?

Program Coordination

Comment on the coordination between the programs being provided by the granting councils and other funding organizations, provinces, and/or amongst themselves. Are there areas for improvement?

One of the challenges our members have encountered with the coordination of programs is around CFI's requirement for matching provincial funds. Research priorities of the federal government (via CFI) and provincial governments often differ, meaning good projects and important national infrastructure may not get funded if the province(s) do not support them. As an example, in the past two years, the BC Knowledge Development Fund has only provided funding for projects that meet the economic priorities of the BC government, i.e. commercialization, technology spinoffs, job creation, etc. Therefore, a number of CFI projects have not received BCKDF

¹ <http://www.sciencereview.ca/eic/site/059.nsf/eng/00023.html>

support because they did not sufficiently meet these economic priorities. There is a real need for greater coordination and communication between federal funding agencies and their provincial counterparts to ensure that projects that have been approved at the federal level are also supported provincially.

Balance Across Funding Elements

Is there a need for the federal government to improve the balance across funding elements (e.g. investments in principal researchers, funding of research staff and other direct costs of research, funding of infrastructure and equipment operations and maintenance, and reimbursement of indirect costs)? If so, how can this balance be achieved? What is the appropriate federal role in supporting infrastructure operating costs? Do CFI and granting councils programs work in a complementary fashion?

With the rapid increase in data intensive research and the trend towards open science, Canada needs to develop a more comprehensive infrastructure to support the data management needs of a growing number of researchers across domains.

The Canadian Association of Research Libraries (CARL) is engaging with a multitude of research stakeholders to help tackle these issues, and is an active participant in the Leadership Council on Digital Infrastructure.

To date, there are no funding mechanisms to support the development of the infrastructure and services needed for research data management beyond the project level. As noted in a 2014 report published by the Tri-Agency, “Currently, research data infrastructure is targeted at the project level, leaving few avenues to support long term access and reuse to the valuable research data produced in Canada. Internationally, Australia, United Kingdom, Netherlands, and the European Union are investing in such horizontal infrastructures to support leading-edge research in those regions.” One consequence of this current approach to infrastructure funding in Canada is that many datasets are no longer accessible and have been lost forever. Research data is an important asset that requires stewardship during and after a research project, and new scientific discoveries and innovation will increasingly rely on the integration of diverse datasets with other types of content, such as publications and other research products. Additionally, proper research data management must be advanced and supported at all levels: the researcher and project, university, as well as regional and national networks. These activities do not come without a cost and we need to ensure that there is funding available for managing data across the entire lifecycle. We believe that a modest (e.g. 2.5%) increase in federal support could be matched by institutions to provide adequate support to research data infrastructure on local campuses.

The overall funding of research in Canada needs to incorporate the costs associated with research data management, discovery and reuse in two significant ways. First, research data management expenses should be an allowable cost that researchers are encouraged to request in their project-level applications. Among such acceptable costs should be a data manager for a project. A general impression now held is that the inclusion of such costs will either be disregarded by funders or be taken from other expense categories, deterring researchers from including data management costs in their applications. Second, the indirect costs of research should include a specific allocation to support data stewardship beyond the project lifecycle and into the long-term stewardship of research data. Internationally, five percent of project costs are seen as appropriate for data management expenses.² These are costs that should be viewed as requiring additional funds and should not be treated as a reallocation from existing sources.

In order to remain competitive, Canada must begin to develop an integrated and comprehensive infrastructure for open science and research data management, including appropriate funding to build horizontal services, not just domain-specific vertical services. CANARIE is an example of horizontal investment in national infrastructure for a high speed network. To that end, in 2014, CARL launched the Portage project to begin to lay a foundation for a national research data management network. The Portage Network is dedicated to the shared stewardship of research data in Canada and is advancing this mission through capacity building and establishing national RDM services and infrastructure. With support from the University of Alberta, Portage has already made available the DMP Assistant, a bilingual, Web-based application that is freely available to all researchers and that provides specialized guidance to help researchers develop their own data management plans.

The Portage Network and partners aim to provide a comprehensive, federated research data service, consisting of platforms that support the stages of the research data lifecycle (planning, ingest, preservation, discovery, access, repurposing), and that offers specialized digital preservation services for research data, including replicated archival storage. These services and supporting infrastructure will be distributed among local, regional, and central nodes and will interoperate through standards and shared protocols. With Compute Canada and other partners, Portage has already developed a proof of concept for a federated research data repository and is collaborating with them in the development of production platforms. Portage will be coordinating services to leverage this new platform and other data repositories as a

² Kathleen Shearer (2015), "Comprehensive Brief on Research Data Management Policies" at http://www.science.gc.ca/default.asp?lang=En&n=1E116DB8-1#_ftnref98. Also Barend Mons (2016), "Science as a Social Machine," International Digital Curation Conference. <https://www.eudat.eu/open-science-as-a-social-machine>

federated network, but further funding is needed to maintain and advance these services to support a greater number of researchers.

To be successful, the organizations making up the wider research stakeholder community must be committed to achieving the national stewardship of digital research data. CARL recognizes this as a shared responsibility across this community. Research libraries have been engaged in local and regional roles in Canada's higher education system, developing a service capacity that undergirds such infrastructure. Our libraries are working with researchers, research service offices, research ethics offices, campus computing services, and senior administrators to establish institutional commitments to this national goal. CARL's members are investing in Portage to help coordinate and facilitate national-level support for local and regional research data activities and to achieve efficiencies through service coordination. As an engaged player in research data management, CARL is open to working with other stakeholders as they too contribute to the stewardship of research data. This includes working with Canada's federal research agencies in their roles to fund the production of research data and to invest in partnerships in the shared stewardship of research data.

Supporting the Needs of Multidisciplinary Research

Are current federal programs supporting the needs of multidisciplinary research programs? If not, how can the situation be improved? Does the funding ecosystem (funding councils and other agencies) work collaboratively and effectively across disciplines?

Although Canada has been trying to encourage more multidisciplinary research, there is still a great deal of emphasis on the individual domain, which is problematic in terms of research data management infrastructure. It creates silos, which limits our ability to integrate and analyze content across domains, including datasets. For example, we cannot analyse the impact of climate change without touching on issues related to agriculture, smart cities, social environments, psychology and so on. Yet, because we still organize our data by domain, we lose opportunities to discover important relationships.

Other regions are developing solutions, such as in Europe where the European Open Science Cloud (EOSC) is in planning. The EOSC will federate existing and emerging horizontal and thematic data infrastructures, in order to bridge today's fragmentation and ad-hoc solutions. They state that it will provide EU researchers an environment with free, open services for data storage, management, analysis and re-use across disciplines. It will add value (scale, data-driven science, interdisciplinary, data to knowledge to innovation) and leverage current and past infrastructure investment (10

billion Euros per year by Member States and two decades European Union investments)³.

In order to remain competitive, Canada must also begin to develop an integrated and comprehensive infrastructure for open science and research data management, including appropriate funding to build horizontal services, not just domain based vertical services, just as we have national infrastructure for a high speed network such as CANARIE.

International Programs

Are there international programs, structures, models, or best practices that Canada should consider adopting? If so, please explain why these should be considered.

Given the current gaps and restrictions for funding research infrastructure in Canada, one model we'd like to point to is the Horizon2020 e-infrastructures program. Through this program, the European Commission has funded the development of leading edge infrastructures including the open access repository network, OpenAIRE and the European research data management network, EUDAT. This program has enabled Europe to develop European infrastructures, bridging the silos across European countries. A similar approach to national infrastructure should be considered in the Canadian context.

In the US, the Andrew W. Mellon Foundation funds projects to develop scholarly communication infrastructure. This type of funding is completely absent in Canada. "The Scholarly Communications program assists research libraries, archives, museums, universities, presses, and arts organizations that seek to realize this potential, and thereby to further our collective understanding of societies and cultures around the world. The Scholarly Communications program promotes the common good by supporting the creation, dissemination, use, and preservation of original sources, interpretive scholarship in the humanities, and other scholarly and artistic materials."⁴.

Other Comments

Are there any other issues or questions that you would like to raise and address?

Another key aspect to creating a more comprehensive and efficient research environment is the use of unique identifiers in maintaining its research records. Related information making up the research record is often scattered across multiple systems, e.g., human resources; grant application systems; research information systems, publications databases; repositories and web pages. The adoption of unique

³ <http://ec.europa.eu/research/openscience/pdf/eosc-workshop-06-2016/eosc.pdf#view=fit&pagemode=none>

⁴ <https://mellon.org/programs/scholarly-communications/>

identifiers for datasets, authors, publications, and institutions will enable the establishment of connections across information systems to trace research outputs having a far greater understanding of the impact of the research Canada funds. CARL strongly urges the Tri-agencies, along with other federal departments that undertake research to encourage, or even require, the use of unique IDs, including ORCID, DataCITE DOIs for datasets, and institutional IDs.

As we move towards this integrated environment whereby we gain a much greater understanding of our research productivity, CARL strongly urges all actors to adopt systems that adhere to the following principles, developed by Science Europe⁵ :

1. Flexibility: research information systems should be flexible enough to allow for extensions in terms of the data objects covered, their definitions, metadata, and use of external data sources.
2. Openness: research information systems data should be available for external use – in line with the principle ‘as open as possible, as closed as necessary’ – and their processing should never require the loss of ownership in underlying raw data by the originating institution.
3. FAIRness: research information systems should foster the findability, accessibility, interoperability, and reusability of the data they store by implementing the FAIR Guiding Principles⁶ for research activity data.
4. Data entry minimisation: research information systems should minimise the need for entering data and facilitate the reuse of data entered manually, in line with the motto ‘enter once, reuse multiple times.’

And finally, we are in an era of unprecedented ‘openness’. Open Science has the potential to transform science through ICT tools, networks and media, to make research more open, global, collaborative, creative and closer to society. “It will fundamentally change the way research is carried out, disseminated, deployed and transformed by digital tools, networks and media. It relies on the combined effects of technological development and cultural change towards collaboration and openness in research”.⁷

⁵ Towards Data Integration for Research Funding and Performing Organisations: a Science Europe Initiative. Science Europe. September 2016 (forthcoming).

⁶ Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*3:160018 doi: 10.1038/sdata.2016.18 (2016).

⁷ <https://ec.europa.eu/digital-single-market/en/open-science>

The federal government and research funders are adopting open access, open data and open science policies and principles. However, open policies and principles require new tools, services and infrastructure. Canada could accelerate our progress towards open science by providing funding mechanisms to support new developments in this area.

For example, robust, comprehensive and economically viable services are already being developed as alternatives to commercial journals. All U15 libraries have repositories that support the Tri-Agencies' open access requirements and can be developed further in a global knowledge commons. Canada is also well positioned with flagship initiatives that can support cooperative sustainable journal publishing models.⁸ The goal is to have academic content in the control of the academy.

⁸ See the CARL white paper for more information: http://www.carl-abrc.ca/wp-content/uploads/2016/04/Can_Univ_Sustainable_Publishing_2016.pdf