

Research Data Canada
Education and Training Subcommittee
Progress Report to the Research Data Canada Steering Committee
January 21, 2014

It is well acknowledged around the world that “digital data are revolutionizing the way research is being carried out, leading to a new data-centric way of thinking. However, as the volume of research data grows exponentially, so must the efforts to ensure that they are preserved, accessible, and understandable.”¹

Research data

Research data as defined by Research Data Canada is “a digital object either produced or managed as a resource for the purposes of evidence in e-research.”² In the context of education and training programs for research data management (RDM), the subcommittee expressed it as any digital information generated through research that is retained for verification or re-use.

Education and training

The terms, ‘education’ and ‘training’ encompass a variety of learning opportunities needed to support a culture of stewardship for Canada’s research data. In higher education, a base level understanding of RDM is as relevant and as important to the education of researchers as matters such as academic integrity and research ethics, and should be woven into the graduate student experience. As well, higher education programs with a RDM focus are needed to produce the highly qualified personnel involved in the broad range of data management activities, from administrative to technical aspects. The term training is used broadly, generally referring to the learning opportunities available to any stakeholders at any stage of their education or career.

Audience

Managing research data requires specialized knowledge and should ideally be supported by skilled data and information professionals throughout the research data lifecycle, as well as a research community that is well educated on the principles of data stewardship. RDM education and training is needed for both the researchers who are producing data (creators) and the managers and staff responsible for research data management services (stewards), as well as those accessing and using data for research (users). These stakeholders are found in multiple sectors: universities, government, health and the private sector. The focus of this report is on education and training for creators and stewards.

Intersecting issues

Policy: Experience in other countries has shown that funding agency policies relating to RDM can be an important incentive for the implementation of RDM education and training

¹ Mapping the Data Landscape: Report of the Canadian Research Data Summit, December 2011, pg. 5

² The Research Data Canada Glossary of Terms, November 2013

programs, just as the policies of universities and employers can play a role. Such policies in Canada need to be updated to include RDM requirements, and in order to adhere to them, stakeholders must have sufficient knowledge of RDM standards, processes, and the research data landscape.

Infrastructure: Research data management also requires an appropriate supporting infrastructure. This includes data repositories that collect, preserve and disseminate research data, together with a high bandwidth research network, and skilled people who provide necessary services. Currently, there are no definitive standards regarding such infrastructure, but internationally there are emerging initiatives such as CARDIO (Collaborative Assessment of Research Data Infrastructure and Objectives)³ and some research funders, such as the UK Engineering and Physical Sciences Research Council,⁴ have statements regarding expectations.

Funding: Any new initiatives in the area of training will come with cost implications. All stakeholder communities, including funding agencies and institutions, must be aware and willing to share the costs of ensuring producers and stewards of research data in Canada have the necessary skills. Research into data management cost modeling exists, but with minimal consideration of training costs (for example, the UK Data Archive Data Management Costing Tool⁵); more work is needed in this area.

Current RDM education and training landscape

In winter/spring 2013, the subcommittee undertook a review of existing training opportunities in the area of RDM for the three major stakeholder communities: creators, stewards and users. The attached table lists the specific initiatives identified through the review. It is not intended to be exhaustive, but rather to give a sense of the range of opportunities available.

Training and education are offered in a variety of different formats:

- ◆ Embedding curricula into graduate programs
- ◆ Semester long university courses
- ◆ Short-term intensive courses
- ◆ Workshops and conference sessions
- ◆ Online training courses

In addition to the formats described above, the UK, which is especially advanced, has the Digital Curation Centre (DCC)⁶, a centre of expertise that offers a suite of training activities as well as other services that assist institutions in implementing RDM programs.

Outside of Canada, there is also considerable attention to efforts to embed RDM in academic curricula. The research projects related to this aim include, for example, the DigCurV⁷

³ <http://cardio.dcc.ac.uk/>

⁴ <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/expectations.aspx>

⁵ <http://www.data-archive.ac.uk/create-manage/planning-for-sharing/costing>

⁶ <http://www.dcc.ac.uk/>

⁷ <http://www.digcurv.gla.ac.uk/index.html>

curriculum framework for digital curators, and the information literacy lens on the Vitae Researcher Development Framework⁸.

Challenges and opportunities in Canada

The final report from the 2011 Research Data Summit identified three goals in the area of capabilities and education for research data management⁹:

- ◆ Data management training modules implemented into research methods courses.
- ◆ Training in data management available to researchers across Canada.
- ◆ Expertise developed to provide support for researchers to help them with their data management efforts.

Currently within Canada there are few training opportunities available in the area of RDM. Canada lacks the national-level coordination of a body like Jisc in the UK, and federal granting councils do not yet provide the necessary policy incentives regarding RDM. We are, however, beginning to see 'bottom-up' development of education and expertise across the country. In the university sector, the Canadian Association of Research Libraries (CARL) is undertaking initiatives to facilitate the development of expertise amongst stewards, through a short RDM services course, a community of practice and most recently a project to develop a library-based RDM network that would include a focus on education and training. In many institutions, individuals are creating education and training programs by drawing on international examples, and they often note that it would be beneficial to have Canadian coordination of both training materials and human expertise to draw upon.

Recommendations

Canada needs a multipronged strategy for the delivery of education in order to build capacity for research data stewardship in Canada. This should include integrating RDM in graduate curricula for future researchers, implementing RDM courses in information schools and other relevant academic programs, and providing a variety of training that will assist current researchers, librarians and other stakeholders to up-skill for RDM. In order to achieve this, we propose the following recommendations as next steps.

1. **Create an entity similar to the Digital Curation Centre (DCC) in the UK that would be a national resource to all sectors.** This would be a coordinating body that would help to facilitate the development of expertise and build upon programs emerging in regions across the country. Given the size of Canada and the tendency towards regional coordination, these services could be spread across several geographically-distributed centres of excellence as suggested by the TC3+ *Capitalizing on Big Data* consultation paper. Such an initiative should be aligned with the RDM services network that CARL envisions in the university realm, helping to coordinate the education and training aspects of this network and broadening it to all sectors (university, government, health, private sector).

8

[http://www.vitae.ac.uk/CMS/files/upload/Vitae Information Literacy Lens on the RDF Apr 2012.pdf](http://www.vitae.ac.uk/CMS/files/upload/Vitae%20Information%20Literacy%20Lens%20on%20the%20RDF%20Apr%202012.pdf)

⁹ Mapping the Data Landscape: Report of the Canadian Research Data Summit, December 2011, pg.10

2. **Adopt a cascading training model in order to maximize the diffusion of knowledge about research data management.** This model focuses on building the skills of trainers and specialists with the intention of training others. For example, training programs could target educating stewards, who would then be tasked with offering training to the current researchers at their institution, who are unlikely to attend more formal training programs. As education in RDM is becoming a programmatic focus in information schools, students in those programs could be drawn upon to teach in doctoral training centres or perform RDM ambassador roles.
3. **Implement training requirements for research data management with stakeholder communities.** In terms of current researchers, one approach has already been successfully implemented in the area of research data ethics in Canada, whereby the completion of a course was required by universities to support the Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans. This approach should be investigated with research data management. In addition, it may also be fruitful to work with graduate programs and professional societies to encourage them to adopt research data management skills criteria into their curricula and certification programs.
4. **Engage internationally, through participation in initiatives such as the Research Data Alliance.** A subset of the subcommittee is currently working on a proposal for an RDA interest group on education and training.

Submitted by the RDC Education and Training Subcommittee

- ◆ Gary Bader, Associate Professor, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto
- ◆ Michelle Brazas, Manager, Knowledge and Research Exchange, Ontario Institute for Cancer Research
- ◆ Joy Davidson, Associate Director, Digital Curation Centre
- ◆ Geoffrey Rockwell, Professor of Philosophy and Humanities Computing, University of Alberta
- ◆ Seamus Ross, Dean and Professor, Faculty of Information, University of Toronto
- ◆ Kathleen Shearer, Consultant (Resource)
- ◆ Ricardo Soriano, Research Quality Coordinator, Office of the Vice-President (Research and International), University of Manitoba
- ◆ Martha Whitehead, Queen's University Librarian and Board Member, Canadian Association of Research Libraries (Chair)

This document was prepared by Education and Training Committee of Research Data Canada. Use is permitted with attribution.