

National Data Science/RDM Training Meeting

Background Document and Notes

*For the meeting held January 22, 2019, 9:00-13:00
CANARIE AKB Boardroom, Suite 500, 45 O'Connor St., Ottawa*

*Final Version, March 7, 2019
Research Data Canada*

DOI: [10.5281/zenodo.2584006](https://doi.org/10.5281/zenodo.2584006)

[National Data Science/RDM Training Meeting](#)

[Background Document and Notes](#)

[Background](#)

[Participants](#)

[Agenda](#)

[Sample Approach](#)

[Resources](#)

[Meeting Notes](#)

[Background](#)

[Roundtable](#)

[Ideas and Approaches](#)

[Next Steps](#)

Background

This document is intended to provide background for a meeting of a select group of RDM organizational representatives, which will be held on January 22nd in Ottawa. The participants were invited based on their current activities in the area of Data Science/RDM training (DS/RDMT), either directly or through their national/international parent organization. The original conversations took place at the 12th RDA Plenary in Gaborone, Botswana, and the January meeting is a follow-up to those discussions.

The desired outcomes for the meeting include:

- a high-level discussion of a national approach to facilitating DS/RDMT;
- identification of next steps to further define an approach and a pilot project;

- identification of possible sources of funding for a pilot.

Notes will be recorded by Laura Gerlitz and distributed after the meeting.

Participants

- Susan Haigh, Executive Director of CARL
- Jeff Moon, Director of the CARL Portage Network
- Lee Wilson, Services Manager, ACENET/Portage
- Ernie Boyko, Chair of Canadian National Committee of CODATA
- Karen Payne, Associate Director for International Technology, WDS-ITO
- Christine Trauttmansdorff, VP Government Relations and Canadian Partnerships, CIGan
- Matthew Lucas, Executive Director Corporate Strategy and Performance, SSHRC
- Kevin Fitzgibbons, Executive Director, Corporate Planning and Policy Division NSERC
- Alison Bourgon, Acting Director General, Science Policy Branch, CIHR
- Kori St-Cyr, Senior Advisor, Science Strategy/Training Strategy Lead, CIHR
- Isabelle Laplante, Bibliothécaire professionnelle, Centre de documentation collégiale
- Zach Hayes, Policy Analyst, Science Programs and Partnerships (Digital Research Infrastructure), ISED
- David Castle, VP Research, University of Victoria/RDC Chair
- Mark Leggott, Executive Director of RDC
- Laura Gerlitz, RDC Intern

Agenda

This Agenda is proposed as a framework for the discussion, and to ensure the discussion will provide some next steps in the allotted time. Coffee and tea will be available, and a lunch will be provided at 12:00.

- 9:00-9:30
 - Background and Goals (Matthew L. and Mark L.)
- 9:30-10:30
 - Interests and Services: brief description of what organizations are doing/planning.
 - SSHRC (Matthew L.)
 - CIHR (Alison B.)
 - NSERC (Kevin F.)
 - CARL/Portage (Susan H./Jeff M.)
 - CODATA/CNC CODATA (Ernie B.)
 - WDS/WDS-ITO (Karen P.)
 - CIGan (Christine T.)
 - Centre de documentation collégiale (Isabelle L.)
 - CANARIE (Kathryn A.)
 - ISED (Zach H.)

- RDC (Mark L./David C.)
- 10:30-11:30
 - Ideas and Approaches
 - Sample Approach (Mark Leggott)
 - Brainstorming
- 11:30-12:00
 - Next Steps (All)

Sample Approach

In the discussions preceding the meeting, a variety of ideas and approaches were raised. This “sample approach” incorporates most of that early discussion.

The goal is to develop a common framework for DS/RDMT in Canada, and one which can be delivered via a “train the trainer” model, with some deployment of core resources (funding and HQP) where needed to help deliver specific training sessions. Delivery would happen in specific regional/organizational/domain contexts, and in such a way as to build expertise across the country. It would also be desirable to have a curriculum that is developed to the point of allowing articulation of academic and/or professional credit by multiple partner institutions. Also desirable is an approach that:

1. facilitates capacity building at all levels of the ecosystem;
2. supports a modular approach to delivering specific content;
3. supports a face-to-face, hands-on delivery;
4. provides ongoing access to a community of practice and learning materials;
5. addresses interests of all sectors (e.g. federal and provincial government agencies, higher education, non-profit sector, etc.);
6. provides a sustainable model after the completion of a Pilot.

A high-level view of possible Pilot tasks follows.

1. Develop Pilot proposal/budget.
2. Identify funding sources.
3. Identify existing curricula/resources.
4. Develop original content where needed.
5. “Canadianization” of existing content.
6. Package content into:
 - a. Core modules
 - b. Sector modules
 - c. Domain modules
7. Articulate of academic/professional credit options.
8. Deliver Pilot Train the Trainer session
9. Deliver Pilot sessions that reflect a diversity of sectors and domains.

10. Create final report and recommendations.

Resources

There are a host of resources for DS/RDM Training, so this is not an exhaustive list. The CARL Portage Training Expert Group maintains a [list of training resources](#), and the [WDS maintains a list of Member training](#) and other recommended programs. Those highlighted here reflect resources intended to be delivered as hands-on workshops, and specific to the organizations represented at the meeting, or that have been raised on early discussions.

1. RDA/CODATA
 - a. The model of the [RDA/CODATA](#) Summer Institute provides a good intersection with international branding/certification.
2. The Carpentries
 - a. The [Carpentries](#) provide a hands-on approach that addresses both the core curriculum and domain-specific interests.

Meeting Notes

Background

1. The conversation around the need for RDM training has been ongoing, most recently at the RDA Plenary in Botswana's, the draft Tri-Agency policy, and the December 2018 Colleges RDM Workshop.
2. The Tri-Agencies (TC3) are seeking opportunities to support/fund initiatives in advance of the release of the final DM policy.
3. Today's meeting is intended as a high level discussion of RDM training from a national perspective and options to establish a proof of concept/pilot.

Roundtable

4. SSHRC: working with NSERC and CIHR on development of the DM policy. Thinking about micro-investments that can spur conversation and get different communities together: series of Connection grants that support knowledge transfer and get communities (disciplines, libraries, etc.) together to talk about research interests, DM initiatives/what DM means to them; looking into interim support/gap funding; wanting to stay engaged and have an ongoing role in capacity building for RDM.
 - a. [Capacity Building Funding](#)
 - b. What kind of concerns were raised in the online feedback for the draft DM policy? Some suggested slightly different approaches to policy; some said TC3 not doing enough; wondering where funding will come from; diversity of perception about where we are as a country on RDM (e.g. decentralized vs centralized); questions

about how policy will speak to privacy issues, indigenous research, intellectual policy.

5. CIHR: working on a broader framework dealing with RDM and furthering RDM discussions with relevant stakeholders; developed learning module (described in the recent RDC Webinar); working with Portage to promote learning tools and help community enhance knowledge in the RDM space; providing funding for embedded training in DS/RDM, and expansion is being considered; looking at ways to provide embedded funding to build data science capacity; recently launched top up funding for funded research projects related to DS/RDM, and which are aligned with FAIR principles; a number of funded initiatives (e.g. SPOR, CLSA, CRDCN) are data-focused and training-focused; thinking about training for peer review capacity-building; considering data-intensive research consortium (leveraging expertise of individuals), and initiative which would allow trainees and PIs to be more aware of RDM and data-intensive research; focusing on pilots first to demonstrate that these are viable options.
 - a. Comment that top up funds are a good idea, highlights DM with respect to rest of research.
6. NSERC: there is a great deal of variety within science disciplines regarding RDM; worked with other agencies to provide ad-hoc funding for organizations like Portage as priorities are clear; interest in developing similar mechanism to SSHRC's using Connect grants. In general TC3 working more closely compared to the past;
 - a. Is NSERC considering any specific/focused projects of RDM flavour?
 - i. Open to developing something; have looked to Portage as primary vehicle within community; reaching out more extensively to organizations like Portage.
7. CARL/Portage: need to address capacity has been recognized at Portage for a while now; there layers to capacity-building, and have been trying to meet these; Training Expert Group has been working for several years, and a number of primers have been released; modules on DMPS and RDM; identifying and adapting training materials for Canada; training workshops available upon request (difficult to offer as the experts are volunteering); contributed to CIHR's RDM modules; reaching other stakeholders by tailoring outreach to community - have been talking with librarians, researchers, government, other funders and partners (e.g. Compute Canada), and institutions; a key talking point is about culture change - socializing RDM is first step.
8. CODATA/CNC CODATA: focus of CODATA has shifted in 10 yrs as research of all kinds has become more data-intensive:
 - a. Challenge Canada has with RDM is similar to what led to creation of the Data Liberation Initiative (DLI). The DLI is researcher-driven; lots of people signing up who didn't have data expertise, so it was important from the beginning to provide train-the-trainer curriculum, bootcamps, annual training in order for universities to establish data services to meet researcher needs.
 - b. Data training schools targeted at early career professionals; might be an opportunity for Canada to adapt RDA-CODATA schools for our own use.

- c. CODATA is interested in meeting capacity-training needs in Canada.
 - d. Summer institutes have the same capacity problem we see elsewhere - can only offer one or two a year.
 - e. How big are summer institutes? About 70 international participants.
 - f. CODATA is interested in getting into the complexity of science disciplines (they vary more than the social sciences).
 - g. There has been concern from science community re: complexities of linking diverse data sets, sharing data sets.
9. WDS/WDS-ITO: Traditionally the program office has been involved in supporting consortium of data distribution centres (international), not exclusively a Canadian org; involved with a number of programs such as CoreTrustSeal (focused on certification of data repositories); Program Office has listing of RDM resources, including training; has an early career scientist position on Council, going through all RDM training initiatives to sort them out; International Technology Office - talking to WDS repository members to find out what their needs are (e.g. IT, training); setup in order to centralize the development of common infrastructure to support members.
- a. There are tiers of people in the research ecosystem - goal is for researchers to focus on their research, data managers can work on RDM.
 - b. NSERC and CORE talked about certification - what does it mean to recognize or certify a repository? Certification is commensurate with an organization's resources; much of the detail is reflected in policy.
10. CIGan: December 7th RDM Workshop at Centennial College got the conversation going for colleges; there is a pool of expertise that has been developed that doesn't exist elsewhere, and that should be drawn upon; one possible source of expertise for delivery of training is CEWIL (Cooperative Education and Work Integrated-Learning Canada) - supports co-op directors in colleges and universities; there are a few colleges which have done more in this area, Centennial - library community has been very integrated with researchers.
- a. Challenges for this community differ from universities - eg, intellectual property, working with the private sector, which is generally linked to their strategic priorities, and responding to local labour market and economics.
 - b. Lack of indirect costs funding is a big issue, so foundational support is needed over and above project funding.
11. Centre de documentation collégiale: A lot of training activity is happening within networks, and the work in layers; Quebec thinks in a national context and delivers programs accordingly.
- a. The dynamic in every college's research centre is different - research centres may not know all of their researchers, relationships may differ.
 - b. Technology transfer offices - don't have the networks to provide umbrella services, but important to work with them.
 - c. Librarians not as integrated with researcher community in Quebec in small colleges.

12. ISED: working on the DRI strategy, which is focused on full DRI ecosystem; DM is part of that, as well as ARC, networking, research software; trying to view DRI holistically and nationally: What is the national layer of the ecosystem and what needs to be done at the national level?
 - a. One major focus is building the capacity that's needed to support new policies, rapidly growing demand for data-intensive research, anything data-related.
 - b. Not interested in a top-down approach, trying to establish frameworks and clarify rules/responsibilities so appropriate people can have appropriate resources: longer term view is to build something sustainable.
 - c. The system is multi-layered: determining proper role in each layer, how they coordinate, and what role would a national layer provide is key. A training strategy would be supportive to allow groups to move the work forward.
13. CANARIE: not typically involved in training; RDM Workshop scheduled for tomorrow is intended to bring projects together to think them on a national level; facilitating the conversation to get groups together and deliver services.
14. RDC: Doesn't deliver programs, facilitates opportunities to get larger group together to develop priorities, policies, etc..

Ideas and Approaches

15. Can we build on different pre-existing approaches to solve the problem of capacity, that is the inability to meet demand for training?
16. One example of a specific outcome: attract summer training school (RDA/CODATA) to Canada for 2020.
17. Option to take a professional certification approach to training, and provide academic/professional credit for students in higher education.
18. The Inter-university Consortium for Political and Social Research (ICPSR) has a summer program that is partly for credit (though it depends on having a university to recognize it); provides DM training.
19. Compute Canada is advanced training on data science topics and intersects with the Carpentries and Digital Humanities Summer Institute (DHSI).
20. The [FutureSkills initiative](#) (federally funded) created to train the Canadian community in the skills needed for new job opportunities. They have run into challenges getting off the ground, but it is a potential opportunity.
 - a. One detail in the FutureSkills report was a chart from the survey on capacity building: there is a clear difference of opinion on the university vs industry driving training.
21. Core principles:
 - a. Focus on train the trainer model - more palatable approach to a national pilot.
 - b. Focus on face-to-face rather than online, as there is already a lot of online training going on.
 - c. May want to focus in on a specific sector.

- d. Training efforts should directly support the implementation of new policy (e.g. TC3 Dm Policy); linked back to the implementation.
 - e. Recognize there is a variety of needs, but initially narrow in on a particular objective.
 - f. Needs to be regionally accessible.
 - g. Initial effort should also be something that isn't disciplinary-focused; discipline agnostic.
22. Needs buy-in from the community; recognize it isn't clear who will fund this in the future; move towards a co-investment model?
- a. DLI model is a co-investment model; each region sends one person covered under DLI funds (travel only covered).
23. Are we talking one kind of undertaking, or something that is more multi-layered?
- a. Sketch national training strategy (long term roadmap) that includes online delivery, out of that choose what is most important for the pilot to undertake.
24. Accreditation and standards - are there any out there for DM to aspire to?
- a. Field is constantly growing and changing, is that an issue?
 - b. Accreditation through institutional context or internationally recognized approach/curriculum/delivery (e.g. RDA/CODATA, Carpentries)
25. Always consider potential pushback from individual researchers.
- a. expectations/level of minimal training for researchers?
 - b. Open data; replace this with FAIR data.
26. Examples of RDM courses in institutions?
- a. Heterogeneity in what is offered across disciplines, institutions.
27. Audience: consider an initial focus on grad students, early career researchers.
- a. Priority group in getting RDM embedded in research community.
28. Drivers and biggest gaps? Where do we need to focus activities?
- a. Broad across disciplines.
 - b. Multi-pronged approach.
 - c. Canadian point of access to provide information; part of this making resources available.
 - i. Can leverage existing opportunities and make them more accessible to researchers.
 - ii. Can also use this activity to find gaps in tools, programs.
29. Opportunity to highlight value of RDM via showing that data basis of AI has to be high quality (e.g. Google Translate's use of an AI engine and training with high-quality datasets like Canadian Hansard.)
30. Universities typically offer core "Research Methods" course in most broad disciplines, there may be an opportunity to have RDM discussions in some of these courses.
31. Long term vs what can we get off the ground in 12 months - possible timing considerations.
- a. Enable universities to address shorter term questions they may get.
 - b. 2-6 months: development of a strategy

- c. 6 months: raining session for development of Institutional Strategy as per the TC3 DM Policy.
 - d. 12 months: offer Carpentries program in suite of pilot institutions.
 - e. 18 months: offer RDA/CODATA summer institute, with goal of enhancing with Canadian content and faculty.
32. Connection grants - where do they fit?
- a. Don't know the answer to this, but they're open to broad ideas.
 - b. Might be possible to pilot Carpentries session with a Connection grant.
 - c. Need to know what is going on with other universities, develop this more efficiently.
33. One potential approach, funding support for specific group that specializes in training (e.g. Carpentries), or is it more of a competitive approach where a group can get national funding?
34. CODATA summer institute - what is the balance of international vs national?
- a. The thinking with this is to have a "Canadianized" curriculum.
35. Access to RDM 101 course for students?
- a. Do this in collaboration with employers? What kinds of skills do they require?
36. Carpentries: great with teaching foundation, no clear on content in the social sciences.
- a. Need to clarify when it came to disciplines if Carpentries have content.
 - b. Possible program is currently developing additional domain-focused training.
37. Coordination: more formally embedded and recognized.
38. Could we create an RDM hub?
- a. Portage has been talking about a researcher dashboard for RDM resources
 - i. Could be a place to collect training materials.
 - ii. Could be focal point for any training sessions.
39. Would be great to simultaneously announce an institutional policy training workshop with TC3 Policy launch.
40. TCPS core module: example of pre-existing required module that involves training.

Next Steps

- 41. Develop details for a 18-24 month support pilot.
- 42. RDC will provide a summary of this meeting (**early February**)
- 43. Create a Terms of Reference for a Working Group
 - a. RDC will create an initial draft TO to send to everyone.
 - i. Draft will be reviewed by Laura, Mark, Lee, Jeff will before going to full group. (**mid February**)
 - ii. TOR will entail further development of Principles.
- 44. Acquire additional details on CODATA/RDA Summer Institute and the Carpentries, and create a profile of what does it means to have the summer school/Carpentries. (**early March**)
 - a. Ernie will cover CODATA/RDA.
 - b. Lee will cover Carpentries.

- c. Develop background information for creating a long term train the trainer made-in-Canada solution, based on these two and others that offer strong options.
- 45. Constitute a subgroup of the Portage Training Expert Group (**April**)
 - a. An immediate vehicle to move elements of the conversation forward.
 - b. Ensure additional representation from groups present at the meeting today.
- 46. Develop pilot proposal for full 18-24 month period. (**May**)
 - a. In the meantime, facilitate the efforts of a specific group (e.g. Portage TEG) that is positioned for training to deliver on the initial TC3-linked effort.
 - b. Assumptions for short-term timing:
 - i. TC3 policy comes out June 2019;
 - ii. Spring 2020: institutions implement requirements;
 - iii. Fall 2020: researcher's submissions have these requirements
- 47. Other meetings where this topic can be discussed, and additional details gathered.
 - a. CODATA meeting in March
 - b. RDA plenary in April
- 48. Other notes
 - a. See agenda for [upcoming UofA workshop](#)