

Research Data Canada response to *Capitalizing on Big Data: Towards a Policy Framework for Advancing Digital Scholarship in Canada*

November, 2013

Research Data Canada (RDC) welcomes the release of the Consultation Document – *Capitalizing on Big Data: Towards a Policy Framework for Advancing Digital Scholarship in Canada* and is pleased to have the opportunity to offer comment.

The RDC Steering Committee members agree that this paper represents an enormous step forward, and one that will advance the digital scholarship agenda in Canada. The paper addresses three of the essential foundational elements that need to be in place in order to move forward – stewardship, coordination of stakeholder engagement, and development of capacity and future funding parameters. These are critical, and rightly recognized as interdependent.

Research Data Canada is an active member of the Digital Infrastructure Leadership Council and has had the opportunity to discuss the document with our sister organizations within the Council. We are aware that these organizations will be making comment independently; indeed, we all have agreed to share drafts of our responses. We also know that the Council itself will be making a submission. In the light of these multiple responses to the consultation document, RDC will confine our comments to those aspects of the document that directly touch our work.

RDC completely supports the premise of the paper that “...research funding agencies, research institutions, and professional scientific associations...should cooperate in the development of clear policies and guidelines....” (Page 1) Currently, digital scholarship is advancing haphazardly without guidelines or policy. No aspect of digital scholarship is more ill served by a policy vacuum than the stewardship of research data.

We applaud the TC3+’s recognition that the landscape involves multiple stakeholders and that there is an extraordinary disposition to collaboration among the players (Page 9). That collaboration has developed uniquely in Canada as a bottom-up activity. In all other jurisdictions that are advanced in building the ecosystem for digital scholarship, advancement has significantly come from government top-down policy direction. Canada’s stakeholders have really pushed collaboration to the frontiers of their capacity until such time as funding and policy direction to harmonize policies across institutions and between institutions and the TC3+ are in place. Coordination is, indeed, essential and should be based on clearly articulated policies and guidelines that span the breadth of activities required to ensure the practice of exemplary digital scholarship in Canada. The consultation document makes reference to

the "bottom-up" actions needed in Canada to shape our research ecosystem and acknowledges the requirement to “balance ... roles and responsibilities among national, provincial and institutional stakeholders ... to ensure both effective support and efficiency (Page 9).” Coordination is vital in bottom-up developments. To realize effective co-ordination, we see the need for new mechanisms to generate collaboration. The scale of national research data management alone requires unprecedented inter-institutional collaboration in building and maintaining research data infrastructure and the engagement of cross-sector support.

We offer Research Data Canada as an example of successful co-ordination. Currently 30+ (see Appendix A) key organizations in the research landscape in Canada have senior staff engaged in the work of RDC either on the steering committee or on the working committees on Policy, Standards, Infrastructure, and Education and Training. Each committee identifies gaps, sources best international practice, and develops recommendations that member organizations can deploy as they discharge their various responsibilities for research data. Research Data Canada is not and never intends to be an operator of research data infrastructure; we exist to bring all the stakeholders together to discover common ground and come to agreement on common practice. It is important to note both with respect to RDC and the wider issues addressed in the paper that co-ordination is not just a compensatory strategy because of an unfortunate plethora of stakeholders. On the contrary, the breadth and variety of stakeholders are essential to developing a truly comprehensive and highly distributed system both for research data and for the entire ecosystem of digital scholarship.

Above all Research Data Canada welcomes the leadership shown by the TC3+ in initiating this consultative process.

We would like to suggest four areas that should be addressed to further strengthen the paper:

1. Long-term data curation
2. The development of data professionals
3. Including within the framework for digital scholarship the data that is generated by government based research and a capacity to accept private research data at the discretion of the company concerned
4. Engagement with the international data community

Long-term data curation

Data needs to be seen both as a product of research and as a foundation for new research. In introductory portions of the paper, this dual role for research data is recognized, but the proposed

actions focus more on plans and capacity for data as the product of original research with a strong emphasis on the responsibility of the researcher. The proposed actions are entirely to be supported, but we would urge that the need for long-term preservation of research data well beyond the period in which the researcher who created them may have interest in them. At some point a service of curation needs to assume the stewardship responsibility of the data from the researcher, and perhaps from the researcher's institution. There are organizations including existing repositories and research libraries with the potential to perform those curatorial services, but none currently have the formal mandate or the funding to provide them.

Development of data professionals

Proposed action 3. c. does acknowledge the need for skills development and graduate and researcher training. We would argue that such skills development within the research community is essential, but it should be acknowledged that we will also require a cohort of data professionals to work with researchers and to perform the curatorial services referenced above. The management of research data requires specialized knowledge and is a highly complex process involving actors from across a number of stakeholder communities. In concert with the development of policies and infrastructure, there is a pressing need to build capacity for managing research data throughout its lifecycle from production to preservation. There are currently very limited opportunities both for education and training for research data management professionals and very limited career opportunities although the need for them is great and growing rapidly.

Cross-sectoral collaboration

Research Data Canada and its precursor the Research Data Strategy Working Group has always emphasized that research happens in multiple sectors, including universities and colleges, government research organizations, and within the private sector and frequently in collaboration among these sectors. A framework for digital scholarship in Canada should acknowledge all research intensive activity in the country and engage all key players. The consultation paper does reference the Government of Canada's Open Government Initiative. We would advocate recognition in the proposed actions of the need for conversation beyond the granting council eligible institutions. It is true that the private sector usually considers its R & D highly proprietary, but voluntary engagement with the wider research community and use of the infrastructure for digital scholarship should be encouraged.

Engagement with the international data community

The consultation paper does speak of the need to understand international best practices. Research Data Canada acknowledges the importance of understanding best practices. We believe we need to do much more. Canada should be actively engaged with international efforts to build research data infrastructure and data interoperability. The entire global community is grappling with many of the same issues. Canada can learn from our peers and also be an active contributor to the development of emerging standards and practices under the auspices of the Research Data Alliance (RDA). The TC3+ should join in the work of the Research Data Alliance Colloquium (RDAC) to fully enable Canadians to build research data interoperability through engagement with RDA. This affiliation will kick start our domestic efforts to build a climate of data stewardship and benefit from the experience of other jurisdictions.

Research Data Canada is grateful for the opportunity to comment on the consultation paper and commits to full engagement in the on-going discussion.

Appendix A: ORGANIZATIONS WITH STAFF ACTIVELY ENGAGED IN THE WORK OF RDC

1. Canada Foundation for Innovation
2. Canadian Association of Research Ethics Boards (CAREB)
3. Canadian Index of Well-Being
4. Canadian Polar Data Network
5. Canadian Research Knowledge Network (CRKN)*
6. Canadian Space Agency
7. CANARIE
8. CARL
9. CASRAI
10. CIHR
11. Compute Canada
12. CUCCIO
13. Department of Energy (US)
14. Digital Curation Centre (UK)
15. Environment Canada
16. McGill University
17. National Research Council
18. National Snow and Ice Data Centre (US)
19. NSERC
20. Ocean Networks Canada
21. OCUL – Scholar’s Portal
22. Ontario Institute for Cancer Research
23. Queen’s University
24. St. Mary’s University
25. SSHRC
26. Statistics Canada
27. Tesera Systems Inc.
28. Treasury Board Secretariat
29. University of Alberta
30. University of Manitoba
31. University of Prince Edward Island
32. University of Toronto
33. University of Victoria

34. University of Waterloo

*CRKN has signaled its intention to join subject to confirmation at a forthcoming board meeting