Response to the

Industry Canada Consultation:  
Developing a Digital Research Infrastructure Strategy

University of Victoria  
September 14, 2015

Contact:  D. Michael Miller, Assoc. Vice-President Research  
Co-chair Research Computing Advisory Committee  
avpr@uvic.ca  250-721-7971

This is a response to the questions posted at

1  How can DRI be realistically transformed, strengthened and supported over the next five years?

A key factor is the provision of long-term stable funding, for infrastructure and for operating costs including human resources. Digital infrastructure has a short life-cycle (less than 5 years). As such, renewal needs to be carried out in a well-planned systematic fashion. Also for institutions to make long term investments and commitments, there needs to be assurance of ongoing support.

Additionally, demand for data storage is rapidly increasing and will do even more so if data curation is to be seriously addressed. The investment needed to respond to data storage requirements, needs to match the projected demands.

The current approach to supporting advanced research computing (ARC) across Canada seems administratively heavy. The reporting pathways through the institutions, regionals (e.g. WestGrid) and Compute Canada to CFI seems cumbersome. It would be prudent to consider whether the processes can be streamlined.

The current model where certain institutions carry the costs of hosting ARC platforms while researchers at other institutions access those platforms free of charge is out of balance. There should be recognition of this in the allocation of operating funds for ARC and DRI.
2 What are the biggest challenges limiting the effectiveness of the DRI ecosystem? What opportunities are there to more efficiently deploy the human, technical and financial resources currently being devoted to DRI? How, and in what priority, should they be addressed?

The existing DRI infrastructure available to researchers in Canada is at least a factor of 10 less than what is needed and data storage is a few orders of magnitude less than what is needed even before the consideration of ongoing data curation.

Human resources are at a premium. This is particularly true for “user interfacing” specialists who can ensure optimal use and access to resources. Funding for such personnel needs to be stable, long-term, and competitive and needs to be considered as an integral part of the overall DRI funding strategy.

The typical model for funding (40% Federal 60% institutional) does not work fairly across Canada if the Provinces do not contribute equitably. This is made worse since the institutions must cover the total funding needs of the regions and central operations.

The current model of planning, procuring resources and allocating resources (through the RAC) focuses on large-scale users (e.g. CANFAR, ATLAS, GENOMICS) etc. As ARC and DRI becomes of greater and greater importance across a wide scope of disciplines, care needs to be taken to properly address the full range of user requirements.

Planning of the infrastructure matches the needs of many “large” users (at least some of them). Some “large” users and the majority of medium and smaller users are not part of the planning of the model. A small user is allocated the default allocation, but the access latency is random and often prohibitive to the research program.

Development of smaller, decentralized, quickly reconfigured, technology tracking resources providing low latency access to medium to small users is becoming critical. Access to nationally funded DRI and ARC should not depend on where a research group is located in Canada and the operating costs should be equitable shared by those who benefit from it.

3 What do you see as the biggest challenges to effective data management and the development of data standards in Canada? What could be done to promote a more rigorous and coordinated data management system that supports research excellence and maximizes the benefits generated by our investments?

Please see the attached response to the Draft Tri-Agency Statement of Principles on Digital Data Management dated September 4, 2015 which we believe addresses key challenges for effective data management.
We emphasize that it is predictability of the existence of resources (infrastructure, operating and human) dedicated to data management issues that will be pivotal to success in this area. Integral is the development of policies, and appropriate infrastructure, that would address data security (when such data requires it), data curation, and data archiving.

It is critical to take into account that a great number of researchers and research groups in Canada undertake collaborative work with researchers, institutions and research organizations in other countries with some initiatives being multi-national. Data collection, management, curation and sharing has to be considered from a multi-national perspective and it is essential that practices in Canada encourage and support collaboration and not introduce impediments.

Data management not only requires the storage and retrieval options of raw data. Large data sets from experiments, observations or simulations require dedicated domain- and subject-specific cyber infrastructure components for meaningful, analytical and contextual data retrieval. A transformed DRI strategy should address the highly-skilled human resource needs associated with these challenges.

4 What is the current capacity within post-secondary institutions to support research data curation?

Again, please see the attached response to the draft Tri-council statement. Researchers do not have the research funding to support data curation. To divert funds for this purpose from existing grants would significantly and negatively affect research programs and graduate student and post-doc support and education. Likewise, institutions do not have funding to support data curation to the extent outlined in the Tri-council draft, either through institutional resources or the Research Support Fund. If data curation is to be seriously addressed it needs to be properly funded separate from current research funding. There should also be a careful assessment how Canada’s interests are best served by a balanced approach to providing support for enhancing research programs and HQP training, and supporting data curation.

5 What are the biggest strengths of the DRI ecosystem? How will these strengths be affected and prioritized by a transformation of DRI in Canada?

The networking infrastructure provided through CANARIE is, and will continue to be critical to DRI across Canada. It is of the utmost importance that the research network keep pace with advances in DRI platforms and use.
Over the past few years various institutions, groups and collaborations have started to accumulate expertise and capabilities that in some instances do an exceptional job. A transformation of DRI should be mindful of such existing initiatives and networks of excellence and innovation in this area, and should build upon them and integrate them, not displace or disrupt them.

6 What is the role of the private sector in supporting a strong DRI ecosystem in Canada?

From the post-secondary institution point of view, the private sector is a key partner in the support of research and particularly in the education of HQP. In the case of HQP, industry contribution through co-op programs and the various opportunities through MITACS and tri-council partnership programs is crucial to ensuring Canada has the highly-skilled talent required to use advanced digital infrastructure for research and other purposes.

Research groups wherever possible use applicable private sector solutions. One goal of the DRI should be to target the areas that the private sector does not cover, e.g. specialized demands or high-risk innovative approaches, and to enable and encourage the adoption of solutions that can involve complimentary approaches and components from both academia and the private-sector.

7 Do you have any other comments or suggestions to support the development of the DRI strategy? (750 words maximum)

It is important to emphasize that in the case of DRI “one size does not fit all.” Some research areas, e.g. particle physics, generate data faster than it can be stored let alone curated. Others, such as research in digital humanities and longitudinal studies of health and aging, will see data curation as critical because of the need for retrospective and long-term analysis of data.

It is important that infrastructure, operating and human resource investments and policy development fully engage post-secondary institutions and key players from across all disciplines.

The development of DRI strategies should take into account the efforts of consortial groups such as the Canadian Association of Research Libraries (CARL) and the Canadian Research Knowledge Network (CRKN) that have significant and positive points of intersection and interest with the organization noted above. These and similar organizations have undertaken significant work that relates to current and future requirements for DRI such as
the CRKN preliminary report on an Integrated Digital Scholarship Ecosystem\textsuperscript{1} and the CARL response\textsuperscript{2} to the Draft Tri-Agency Statement of Principles on Digital Data Management.

An important role of a DRI strategy should be to support the operation and effectiveness of team science. It has been recognized elsewhere\textsuperscript{3} that team science has become over the past decades an increasingly important mode of research. The largest, most complex and highest impact research activities are now pursued in moderately-sized or large teams that face numerous challenges, such as geographical distribution, knowledge and capabilities integration, as well as transparent access to and sharing of data products, analytics and representations. A DRI strategy should allow for flexible components that can address the multiple, diverse and evolving needs of team science nationally and internationally.

\textsuperscript{1} http://crkn.ca/node/1347
\textsuperscript{2} http://www.carl-abrc.ca/uploads/pdfs/Public\%20Policy/Portage\%20response\%20to\%20Statement\%20on\%20DDM-FINALe.pdf
\textsuperscript{3} http://www.nap.edu/catalog/19007/enhancing-the-effectiveness-of-team-science
Response to the

Draft Tri-Agency Statement of Principles on Digital Data Management

University of Victoria
September 4, 2015

Contact: D. Michael Miller, Assoc. Vice-President Research
avpr@uvic.ca 250-721-7971

The University of Victoria appreciates the opportunity to comment on the Draft Tri-Agency Statement of Principles on Digital Data Management. In our view, there are significant issues to be considered before the Statement is finalized and we look forward to contributing to the ongoing consultation with institutions.

We agree with the general principle of making the results of publicly funded research as accessible as possible. The University also supports access to the data supporting publicly funded research while recognizing there are financial, practical and other constraints that legitimately limit such access. In particular, we are concerned with the lack of clarity regarding funding the substantial costs of creating and maintaining data repositories. It is clear that the current level of Tri-council funding does not allow researchers to contribute to such costs nor does the funding available to institutions through the Research Support Fund. It is notable that the draft document does not address funding agency responsibility for financially supporting data repositories,


1. Expectations

1.1 Data Management Planning – We agree that

   Data management planning is necessary at all stages of the research project lifecycle, from design and inception to completion.

   It is a key aspect of the proper conduct of a research project.

1.2 Constraints and Obligations – It is clear that

   Research data must be managed in conformity with all commercial, legal and ethical obligations.

   In particular, the University is committed to ensuing research adheres to the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – 2nd edition, and the Tri-Agency Framework: Responsible Conduct of Research.
1.3 Adherence to Standards – Standards when available should be adhered to but it is the nature of research to expect many situations where an applicable standard has yet to be developed and standards evolve with ongoing research in many areas.

1.4 Collection and Storage – The second half of the statement

*Data should be collected and stored throughout the research project using software and formats that ensure secure storage and enable preservation of and access to the data well beyond the duration of the research project.*

is very open-ended. How long is “well beyond.” There are numerous examples of out-dated storage media. There needs to be reasonable limitations on the requirement to migrate data to new formats to keep data accessible due to the complexity, cost and time commitments associated with such migrations.

1.5 Metadata – The proper use of metadata is also a key aspect of the proper conduct of a research project and a critical aspect of discovery and long term preservation. Researchers will require more guidance regarding *community best practice*, because some disciplines have well-developed standards, while others have no metadata standards at all. *Community* is a somewhat vague term: is that discipline, institution, or science writ large? Does it include library metadata standards?

1.6 Preservation, Retention and Sharing – The statement

*All research data resulting from agency funding should normally be preserved in a publicly accessible, secure and curated repository or other platform for discovery and reuse by others.*

raises the questions of what is the normal case and what is not and who makes the decision. There is also the very significant question of how long data must be preserved and who will cover the cost of preservation. The draft puts the onus on the researcher to determine whether data needs to be preserved. What is the outcome if a researcher makes a good faith decision but others disagree and in particular claim the required data has not been retained?

1.7 Timelines – There are certainly circumstances where a researcher may have legitimate reasons for not making data public. Regarding the issue of *a defined period of exclusive use of data for primary research* it is not clear who will determine what a reasonable period is and in which cases this applies. It the researcher makes this determination, the issue again arises as to what are the ramifications of a decision others do not agree with.

1.8 Acknowledgement and Citation – This is clear and obvious proper professional behaviour.

1.9 Efficient and Cost Effective – These are obviously key issues. It is not clear who will make these determinations and whether it is the institutions, the researchers or the Tri-Councils, through the Research Development Funds who will cover the costs. Careful consideration is needed to policy and practice in this regard to ensure a level playing field across each institution and indeed across all institutions.
2. Responsibilities

2.1 Researchers

- Incorporating data management best practices into their research, and developing data management plans to guide the responsible collection, formatting, preservation and sharing of their data throughout the entire lifecycle of a research project and beyond;

  It is reasonable to expect researchers to prepare data management plans and follow best practices. As noted above there will have to be clear statements as to what data must be preserved and for how long.

- Following the requirements of applicable institutional policies and professional or disciplinary standards; Agreed.

- Acknowledge and cite datasets that contribute to their research; Agreed.

- Staying abreast of standards and expectations of their disciplinary community. Agreed.

2.2 Research Communities

- Developing data management standards, or promoting existing standards, and working collaboratively to review and improve these standards; No comment.

- Recognizing data as an important research output and fostering excellence in data management within their research community; No comment.

- Identifying and encouraging the use of specific repositories and platforms.

  It is not clear what a researcher’s or an institution’s responsibility is here. If a research community encourages the use of certain repositories or platforms, what is the obligation to follow that advice particularly if the costs are prohibitive? Also, provincial law is quite different regarding the retention and storage of personnel information. In BC, such data must be stored in Canada. This could mean that some approaches are not uniformly applicable across Canada and care must be taken that researchers in one location are not disadvantaged compared to others elsewhere in Canada.

  Further, as noted above, the cost of creating and maintaining data repositories has to be carefully considered. It is not appropriate for a community or funding agency to encourage the use of a repository or platform while simply off-loading the cost onto researchers and institutions.

2.3 Research Institutions

- Providing their researchers with an environment that enables world class data stewardship practices, as well as delivering, or supporting access to, repositories or
other platforms that securely preserve, curate and provide continued access to research data;

It is not clear what “world class” data stewardship practices are. Without better guidance as to what data needs to be in a repository, how long the data needs to be retained and how broad the access is to be, it is impossible to even estimate of providing the environment suggested. Without putting reasonable limits on the expectations, this could very significantly increase the indirect cost of doing research putting excessive pressure on research grant funds, institutional budgets and the Research Support Fund.

- Monitoring compliance of researcher data management practices with legal, ethical and commercial considerations in the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans – 2nd edition, the Tri-Agency Framework: Responsible Conduct of Research, and other relevant policies;

- Providing their affiliated researchers with guidance to properly manage their data in accordance with the principles outlined above and community best practice, including the development of data management plans; Agreed.

- Recognizing data as an important research output and fostering excellence in data management; Agreed.

- Promoting the importance of data management to researchers, staff and students; Agreed.

- Developing their own data management policies and ensuring that these policies are in accordance with the principles outlined above, provincial and national laws, and can accommodate the rapidly evolving research communities’ best practices.

The University has a number of information management policies in place. We will review those and ensure the principles outlined in the final Tri-Agency Statement of Principles on Digital Data Management are properly addressed. If needed the University policies will be amended or new policies put in place. Since this will be required of all institutions, the clearer the final Tri-Council policy is on institutional responsibilities the better. Those responsibilities should be finalized in close collaboration with the institutions to ensure they are workable and cost-effective.

2.4 Research Funders

As noted above, we are concerned not to see funding of the creation, maintenance and use of data repositories and platforms as a responsibility of research funders. The costs cannot be
funded from existing research funding without significantly affecting the supported research programs, and the costs cannot simply be downloaded to institutions.

- Developing policies and requirements that facilitate and recognize responsible data management, in accordance with the principles outlined above;

This should be done in close collaboration with research institutions to ensure funder and institutional policies are consistent and complimentary.

- Providing applicants with clear information and guidance with regard to fulfilling data management requirements; Agreed.

- Recognizing data as an important research output; Agreed.

- Promoting the importance of excellent data management; Agreed.

- Where appropriate, providing peer reviewers with guidance and developing assessment material for including data management considerations in the application assessment process.

It is crucial that applicants be advised of this guidance and the assessment processes to be used so that they can provide informative applications in full knowledge of how the applications are to be assessed.

3. General

It is not clear how the proposed principles apply for a research project funded from multiple sources. Care must be taken to ensure that application of the Tri-Council principles with respect to data collection, retention and sharing do not preclude co-funding from other sources, e.g. industry or from other countries. In addition for international research, there needs to be some flexibility in the application of the Tri-Council principles to be in compliance with policies that might apply to the international partners.