

McMaster University's Institutional Response to
New Digital Research Infrastructure Organization (NDRIO): Call for White Papers on Canada's
Future DRI Ecosystem

Prepared by the Office of the Vice-President, Research with input from the Associate Deans,
Research

OVERVIEW

Canada's current digital research infrastructure (DRI) offerings are in short supply, though the available resources are generally effective. In addition to increasing the availability (scale) of resources, it is critical that researchers be made aware of the digital tools and technologies available to them and that training is provided to ensure researchers and their students are able to make effective use of these tools. People are a critical component of Canada's DRI ecosystem, and a robust model that includes longer-term financial stability for positions that deliver national services is essential, moving forwards.

Research Data Management (RDM) is a fundamental element of modern research so that data is findable, accessible, and reusable. The current Portage Data Management Plan (DMP) Assistant is an effective tool to support funders' desire to increase the efficiency and accessibility of research data. Although tools like the DMP Assistant and exemplar data management plans are useful, RDM experts, resources, and support are also needed to guide and advise researchers to develop their data management plans. Data repositories and discovery platforms that assign persistent identifiers (PIDs) to digital research assets enable Canadian researchers to steward, archive, and curate research data for discovery and reuse. Repositories and discovery platforms with the aforementioned characteristics are important components in assuring the full value of research projects' funding is derived by the public. Repositories such as Dataverse and the Federated Research Data Repository (FRDR) need to be instantiated, operated, maintained, and grown according to demand so that the value of the research results can be fully appreciated.

Support from research data managers, data archivists, and metadata experts is crucial for the effective and efficient use of these systems, in addition to IT support for appropriate management of the underlying infrastructure. Highly Qualified Personnel (HQP) are critical in allowing researchers to concentrate on their projects. A graduate student, for instance, should be concentrating on their field of study while gaining mastery of research tools including ARC, data repositories, and discovery platforms applicable to their work; those same students should not be required to become part-time, under-qualified systems administrators just to accomplish their research. Thus, resources and funding dedicated to supporting and retaining HQP whose sole focus is system administration should be one of NDRIO's focuses moving forward.

As Canada integrates the delivery of all aspects of DRI, the consideration of open science is encouraged. The Government of Canada released a Roadmap for Open Science in February of 2020. As an element of open scholarship, the DRI environment must be designed in alignment with the principles and guidelines outlined within this roadmap. Notably, the principles included in this roadmap emphasize the need for non-commercial archives to enable wide-spread accessibility and ensure all researchers have access to published data.

MCMASTER UNIVERSITY

McMaster University researchers utilize a number of national resources to support their projects. This includes the Graham and Niagara compute clusters, the Arbutus cloud, digital repositories like Dataverse, the Federated Research Data repository (FRDR), and software tools like the Portage DMP assistant. Researchers also use a number of local resources including specialized data processing laboratories, local HPC experts and computational resources, as

well as a number of services provided by the McMaster University Libraries and IT partners across campus.

Despite using these resources, many gaps persist in the digital tools and technologies researchers need to support their research activities. McMaster has undertaken a number of needs assessments related to DRI over the last few years, and common themes have emerged which we share in the hope that they will inform NDRIO’s actions moving forward.

| | |
|--|---|
| <p>Dissemination of DRI knowledge & resources</p> | <p>Many researchers are unaware of what resources are available to them via the national computing platform and how they can take advantage of these. Dissemination of information regarding DRI resources and training opportunities available to researchers in Canada must be considered moving forward.</p> |
| <p>TRI-Council & DRI alignment</p> | <p>There should be alignment between Tri-Council funding and the allocation of computational resources, such that researchers who receive funding are also given access to the resources required.</p> |
| <p>Greater allocation of resources for moderate ARC use</p> | <p>There are a number of researchers that exist at our institution who require more ARC resources than are available through Compute Canada’s base allocation, but who are not competitive enough for dedicated allocation. Compute Canada currently serves “light” and “heavy” ARC users reasonably well; however, “moderate” ARC users are not able to access enough computational resources to meet their needs. The new national model must consider these researchers in ARC resource provision.</p> |
| <p>GPU Availability</p> | <p>Many researchers at our institution require access to GPUs but are not affiliated with one of the AI institutes. These should be available through the national platform.</p> |
| <p>Sensitive data</p> | <p>Our researchers work with a large amount of health-related and sensitive data that have requirements around location. (e.g., some data needs to remain within the campus boundary for a variety of reasons). The new model needs to accommodate location requirements to</p> |

| | |
|---|---|
| | <p>understand that integration with the national platform is ideal.</p> |
| <p>Research Data Management & Research Software training and support</p> | <p>Many researchers struggle with RDM and need access to tools, repositories, and people to help them. Similarly, many researchers need support with research software and data visualization. Training opportunities and dedicated highly qualified personnel support in these areas should be considered in the new national model.</p> |
| <p>Dedicated highly qualified personnel</p> | <p>Highly qualified personnel (HQP) whose focus is on the system administration and infrastructure are critical in allowing researchers to concentrate on their projects. Resources and funding dedicated to supporting and retaining HQP should be one of NDRIO's focuses moving forward.</p> |
| <p>Integration of campus & National DRI</p> | <p>Better integration between campus IT resources and the new national research computing platform will make it easier for researchers to access the resources they require.</p> |
| <p>Continuing with the positive</p> | <p>Although there are improvements that need to be made to the current Compute Canada model, there are a number of positive aspects to the said model that work well for Canadian researchers. There is a wealth of experience within the existing Compute Canada federation. We hope that NDRIO retains and builds upon the knowledge that already exists.</p> |