



**Response to NDRIO Call for  
White Papers on Canada's Future  
DRI Ecosystem**

**December 14, 2020**

---

Martha Crago  
Vice-Principal (Research and Innovation)  
McGill University

## **McGill University**

### **– Response to NDRIO Call for White Papers on Canada’s Future DRI Ecosystem**

The Canadian Digital Research Infrastructure (DRI) landscape is being redefined by the Federal Government to better serve Canada’s research community and support scientific excellence through an important investment in the New Digital Research Infrastructure Organization (NDRIO). McGill University welcomes this opportunity to provide feedback and contribute to inform NDRIO’s Strategic Plan.

#### **Meeting the increased need for Digital Research Infrastructure**

Digital infrastructure is an essential tool for researchers. The need for digital infrastructure has increased significantly over the past few years, which is demonstrated in grant applications where budget requests for computing capacity, storage, applications, software developers, have increased. The optimal return on investments in digital infrastructure is not possible without a national strategy for coordination of efforts. This is especially the case for major funding agencies: CFI in regard to infrastructure investment and with the Tri-Agency and Genome Canada for streamlined allocation of computing capacity and funding for data curation as part of research grant applications.

#### **Lower the barriers for access**

Regardless of the research discipline, scope and scale of the research activities, or technological proficiency, access to digital infrastructure has become a fundamental need. Therefore, removing perceived and real barriers to utilizing DRI resources should be one of NDRIO’s top priorities. Support staff at the local, institutional level are effective ways of providing training and assist with accessing the infrastructure and migrating workflows to a national platform for Advanced Research Computing (ARC). Nationwide coordination of user support and training will assure that the service offerings, and pools of expertise keep in pace with researchers’ needs.

#### **Enable international research networks**

Large-scale research initiatives in national priority areas require high quality computational and data resources that are readily available. Such research initiatives often involve collaboration with research groups in other countries and a desire to exchange research data across jurisdictions. By providing a sustainable and distributed structure for handling data and data retrieval tools to the Canadian research groups and institutions, NDRIO can make a major difference in advancing international collaborations. An example of this would be the provision of funding to research groups to participate in efforts related to system interoperability and data standardization that may be discipline-specific and support universities to establish guidelines and directives related to cybersecurity and privacy as well as legal and ethical frameworks for data transfer and sharing.

## **Advanced Research Computing**

The current model of the national ARC platform under Compute Canada Federation (CCF) has both advantages and shortfalls. Elements to keep and strengthen are training for new users, coordination of capacity allocations, standardization and streamlining of operations and technical support to researchers. This federated model also has challenges. Specifically, the current funding model and decision-making are convoluted and sub-optimized, and a stronger framework is needed describing roles and responsibilities. The current model of host sites operated by universities could be retained and potentially expanded to a larger network of sites, where the sites receive stable and predictable funding for operating and a clear mandate to deliver services that are consistent across the national platform. Scientific users should participate in development of services and oversight of the platform. Complementary services and special expertise at the site level should be encouraged under the umbrella of a national platform that is scalable and adaptable to meet scientific needs.

## **Research Data Management**

Research Data Management (RDM) in Canada has been gaining traction in recent years through significant efforts of stakeholders like the Canadian Association of Research Libraries (CARL), Portage Network, and Research Data Canada (RDC). National research data repositories have incentivized data sharing practices. As an example, Scholars Portal Dataverse has facilitated data sharing across disciplines through its ease of use. The recent Federated Research Data Repository (FRDR), which is still in development, aims to provide important services in a single location: Storage, Curation, Discovery, and Preservation. NDRIO will be uniquely positioned to support curation and coding of research data as part of research funding.

With increased awareness of data management and the forthcoming Tri-Council Policy on Research Data Management, researchers will need training, guidance, and support by professional data experts such as Librarians and Data Managers at the institutions to navigate the complicated landscape of standards for (meta)-data, data repositories, and field-specific regulations including compliance matters on privacy, security and ethical issues. Therefore, funding for institutions to build the human infrastructure is essential. This cannot be achieved without a comprehensive effort across universities. Coordination between institutions with varying degrees of RDM services and support will lead to promoting equitable access to knowledge and expertise. Promoting services for all is of little value if knowledge on how to access them is only available to few.

## **Health data, privacy and security**

Secure data solutions are lagging in Canada compared with other countries, and the sharing of sensitive data, such as health research data, remains a major challenge. With one of the world's richest health datasets, the lack of a secure pan-Canadian DRI system to host and analyze sensitive data limits our ability to leverage these resources into timely and impactful scientific breakthroughs. Linking health data across Canada is difficult due to provincial differences in privacy regulations, making workflows using health data cumbersome and often costly to administer. A federated pan-Canadian DRI system would allow access to sensitive data for research while remaining compliant with institutional policies and provincial regulations, as such data would remain locally. Efforts are already underway and should be supported, such as Canadian Distributed Infrastructure for Genomics (CanDIG) and the Health Data Research Network. This past year has demonstrated that collaborative-interdisciplinary-research can improve our ability to understand and respond to novel diseases, but also identify solutions to old ones.

Establishing secured ARC and RDM infrastructure for health-related research would open unprecedented opportunities and facilitate the collaboration of health professionals, academics, and government scientists.

## **Research Software**

As Research Software (RS) has become an indispensable tool in modern science, funding for the development of specialized software for numerical modelling, data analysis, and visualization has been increasingly allocated in sponsored research activities. Indeed, most research today would be impossible without specialized software. Software is an asset that requires the use of best practices in design, implementation and long-term maintenance to ensure a reasonable degree of quality, reusability and sustainability. Poorly implemented software can not only jeopardize scientific results validity and reproducibility but also expose entire organizations to cybercrime.

Large research groups may have the capacity to recruit professional software engineers to develop specialized research software, but these tasks are often delegated to research trainees, such as postdoctoral researchers and graduate students who are primarily self-taught and lack exposure and incentives to adopt software development practices that are widespread in the broader field of software engineering. There is thus significant demand for students to be trained in proper research software development, but also for readily available software engineering experts to support large-scale initiatives. Local support staff at the institutional level will provide critical support to researchers, students, and researcher personnel for software development. We encourage NDRIIO to implement a RS strategy that incorporates additional focus on training and adoption of best practices as well as promoting a culture of software sharing and reuse.

### **Summary of the key points:**

- **Canada needs a new national platform for digital infrastructure beyond high performance computing that also includes software resources, analytical tools and support for data management**
- **Coordinate funding and resource allocation of digital research infrastructure with the major research funding agencies**
- **Local support with sustained focus on training will expand access to the national digital research infrastructure system**
- **Stable and predictable funding to host sites for operations of the national Advanced Research Computing platform**
- **Adopt a Research Software strategy that incorporates additional focus on training and adoption of best practices as well as promoting a culture of software sharing and reuse**