September, 2021

Meeting the Digital Research Infrastructure Needs of the Canadian Research Community

Digital Research Alliance of Canada | Alliance de recherche numérique du Canada

Funded by the Government of Canada | Canada

Executive Summary

Research in Canada is being transformed by the expanding scope and availability of digital research infrastructure (DRI): the dynamic network of hardware, software, people, organizations, methods, practices, and services involved in creating and mobilizing knowledge. DRI is a catalyst for discovery and innovation, and is critical to ensuring Canada's economic and social well-being.

The Digital Research Alliance of Canada (the Alliance) has been established by Innovation, Science and Economic Development Canada (ISED) to provide Canadian scientists and scholars with the digital infrastructure needed to conduct innovative and world-leading research. Toward the creation of a financially sustainable DRI ecosystem, the Alliance Researcher Council together with the Alliance's Senior Analysts summarized the priorities of the Canadian research community using a variety of stakeholder inputs (current state papers, position papers, national survey and Town Hall meetings).

This document, *Meeting the Digital Research Infrastructure Needs of the Canadian Research Community* is delivered to the Alliance Board of Directors by the Researcher Council in the spirit of collaboration. Its vision of the new DRI system includes a spectrum of accessible services that support the breadth of Canada's research disciplines and computational approaches, prioritizes Indigenous self-determination and data sovereignty, coordinates a robust national program of DRI education and training, and apportions resources transparently and equitably.

The high-level themes and specific priorities are summarized on the following two pages.

National infrastructure

- Immediately invest in additional traditional and cloud computing, expand storage resources, and develop a sustainable financial plan for its maintenance and continual renewal.
- Build a national strategy for the use of in-house and commercial computing clouds.
- Develop and implement a plan for long-term storage, data curation, and data preservation.

Self-determination and data sovereignty for Indigenous Peoples in Canada

- Provide funding support, expertise, and partnership to Indigenous Peoples in Canada so that they can develop their own sovereign digital research infrastructures and use digital methods in support of cultural revitalization.
- Facilitate opportunities for respectful collaboration, which respond to and support community needs, goals, and aspirations.

Equity, diversity and inclusion

- Integrate EDI principles into all decisions and activities as a foundational principle.
- Adopt metrics for EDI that follow national and international standards and best practices, identify areas requiring improvement, address gaps promptly, and engage in continuous monitoring and improvement.
- Provide services and support to all researchers in French, English, and other languages across Canada.
- Ensure all researchers, including those from marginalized groups, have access to the Alliance's tools and services, regardless of geography, discipline, or institutional affiliation.

Professional support personnel

- Ensure that there are sufficient professional support personnel to provide the necessary foundational support to meet the diverse needs of the research community.
- Engage with the regional consortia, sites, and other organizations to put in place a model for the ongoing support and retention of the professional support personnel, and establish a system for the training and education of students and early career staff.

Research data management and stewardship practices

- Develop a strategy and plan for managing sensitive data, and collaborate with other national and regional groups active in this area, including Indigenous communities.

- Build infrastructure and support for the FAIR management, curation and preservation of research data in collaboration with Canadian universities and laboratories, respecting also the Global Indigenous Data Alliance's CARE Principles.

Education and training

- Develop a DRI training curriculum in partnership with Canadian universities that leverages existing programs, and is based on the priorities and expertise of the research community.
- Ensure that documentation is current, well-curated, and user-friendly, and that training programs are easily accessible and offered frequently.

Research software

- Establish support and ongoing maintenance of critical infrastructure software components.
- Review the research software needs of the community to determine the key open-source packages that need to be developed and maintained by the Alliance's professional support personnel.
- Develop a strategy for supporting the development of new and innovative open-source research software.

National and international engagement

- Establish relationships and partnerships with existing Canadian institutions, laboratories, and industry that will help the Alliance define its role in the DRI ecosystem.
- Build connections with the international community that will help the Alliance develop new partnerships and support projects with a shared infrastructure.

Contents

1 Overview

Research in Canada is being transformed by the expanding scope and availability of digital research infrastructure (DRI): the dynamic network of hardware, software, people, organizations, methods, practices, and services involved in creating and mobilizing knowledge. Advances in DRI are broadly impacting all areas of research and education. DRI is a catalyst for discovery and innovation, and is critical to ensuring Canada's economic and social well-being.

In 2018, the Government of Canada earmarked funds for a national DRI strategy to provide Canadian scientists and scholars with the digital infrastructure needed to conduct innovative and world-leading research. In 2019, Innovation, Science and Economic Development Canada (ISED) mandated the creation of the Digital Research Alliance of Canada (the Alliance)[1] to meet these goals.

As a foundational step, the Alliance established a Researcher Council[2] to formalize the regular consultation of researchers in Canada. Representing a range of cultures, languages, identities, expertise, and regions, as well as a broad array of academic disciplines and institutions, the Researcher Council provides advice to the Alliance Management and Board of Directors. Specifically, in 2021, its role is to highlight the needs and priorities critical to Canadian researchers for the development of the Alliance's strategic plan and to help determine the investments for the first years of operation.

In 2020-2021, the Senior Analysts of the Alliance, with the guidance of the Researcher Council, organized four consultative initiatives to understand the current state of DRI and researcher needs in Canada:

- **Three current state papers** focused on advanced research computing (ARC), research software (RS), and the research data management (RDM) ecosystems in Canada, led by the Alliance's Senior Analysts with the input from subject matter experts, including researchers from within and beyond the Researcher Council.
- **107 position papers[3]** were submitted by Canadian researchers and research institutions, describing their DRI requirements, challenges, and recommendations.
- A **national survey** collected information from 1381 researchers across Canada, including information on demographics and academic disciplines.
- A series of national **Town Hall meetings** in French and English was attended by 477 researchers and other stakeholders in the DRI community who helped validate the collective findings from the community consultations.

[1] Digital Research Alliance of Canada (the Alliance). https://alliancecan.ca

[2] The Alliance Researcher Council is a group of researchers, selected from across Canada, for their expertise in DRI. https://alliancecan.ca/researcher-council

[3] https://alliancecan.ca/canadian-digital-research-infrastructure-needs-assessment/position-paper-submissions

World-class and innovative research requires leading-edge infrastructure to collect, store, analyze, and share research data and results. DRI is used in all academic research from providing access to astronomical images, to creating artificial intelligence (AI) algorithms, linking diverse health data, and interpreting data in the social sciences and humanities.

The Alliance has been established as the architect of Canada's DRI landscape; its strategic plan will guide activities to 2024 and beyond. However, the Alliance will not be able to deliver its core mandate alone; to succeed it will need to collaborate extensively with institutions, provincial and regional bodies, and national laboratories and organizations to develop a coordinated approach for delivering a highly complex suite of services.

The Alliance's *Guiding Principles* describe how the organization aims to be researcher-centric, service-oriented, accountable and transparent, striving for excellence, and collaborative.[4] These principles must be woven into the Alliance's organizational structure, and all its activities and initiatives. Collaborative relationships, based on a clear division of roles and responsibilities, will be key to moving forward. Investing in infrastructure and professional support personnel will be critical to meeting the needs of Canada's broad and diverse research community.

This document is written by the Alliance Researcher Council, together with the Alliance's Senior Analysts. The group has significant expertise in all areas of DRI and is committed to the long-term success of the Alliance.

The needs of the Canadian research community identified through the consultations are organized here into themes that highlight the key priorities for the Alliance from the perspective of the Researcher Council:

- National infrastructure
- Self-determination and data sovereignty for Indigenous Peoples in Canada
- Equity, diversity and inclusion
- Professional support personnel
- Research data management and stewardship practices
- Education and training
- Research software
- National and international engagement

The Researcher Council submits this document to the Alliance Board of Directors and Management in the spirit of collaboration for the creation of a financially sustainable DRI ecosystem. This vision of the new DRI system includes a spectrum of accessible services that support the breadth of Canada's research disciplines and computational approaches, prioritizes

[4] The Alliance Guiding Principles. https://alliancecan.ca/about/guiding-principles

Indigenous self-determination and data sovereignty, coordinates a robust national program of DRI education and training, and apportions resources transparently and equitably.

2 National infrastructure

The Alliance will provide a national infrastructure of hardware, software, and personnel that will serve as the foundation for specialized platforms, data management, and computing systems, among other services.

Compute Canada[5] currently provides advanced research computing (ARC) resources to researchers, with two thirds being satisfied with their provision according to the Survey results. In order for Canada to reach the average of the G7 nations in terms of compute per Gross Domestic Product (GDP), Canada will need to double its computing resources. The Alliance should set a goal to reach this milestone by the end of its first full mandate.

The need for additional computing capacity is being partially addressed by the increasing use of cloud computing. Compute Canada offers an in-house compute cloud for its users while a number of researchers are purchasing commercial cloud services for compute-intensive applications, or for access to unique or specialized platforms and software. While commercial clouds are still considered more expensive than in-house clouds for compute-intensive applications, they can be cost-effective if managed in an opportunistic manner. To support growth in demand, the Alliance should develop a national strategy for cloud computing that includes in-house clouds and explores partnerships with commercial service providers, while ensuring data ownership, security and academic freedom.

The current level of data storage provided by Compute Canada meets the needs of approximately 60% of researchers according to the Survey results. Currently, Compute Canada does not provide the long-term data and archival storage that is crucial for improving replicability and reusability, and embracing open research and data principles. The Alliance needs to establish long-term data storage systems to address these needs, and to make more efficient and effective use of its existing storage resources.

From an operational perspective, researchers require reliable and stable infrastructure that is available even during scheduled maintenance periods. The Alliance should identify and resource appropriately the services that require near 24/7 levels of service operability.

Cybersecurity is becoming a critical issue for DRI ecosystems worldwide. The Alliance needs to ensure that its systems and the data collected by researchers are protected. The Alliance's Canadian DRI partner, CANARIE[6], was mandated by ISED to take a leading role in cybersecurity. Together, CANARIE and the Alliance should offer cybersecurity guidance, and services, as well

[5] Compute Canada. https://www.computecanada.ca

[6] CANARIE, Canada's National Research and Education Network Organization. https://www.canarie.ca

as sharing guidelines and threat-awareness training, along the lines of the Ontario Cybersecurity Higher Education Consortium model[7].

Priorities

- Immediately invest in additional traditional and cloud computing, expand storage resources, and develop a sustainable financial plan for its maintenance and continual renewal.
- Build a national strategy for the use of in-house and commercial computing clouds.
- Develop and implement a plan for long-term storage, data curation, and data preservation.

3 Self-determination and data sovereignty for Indigenous Peoples in Canada

Recognizing the historical and ongoing unethical uses of Indigenous data within the context of Canadian-led research, the Alliance must ensure that First Nations, Métis, Inuit, and other Indigenous researchers and communities can fully participate in the development of the Alliance and the identification of its priorities. This will require that the Alliance's policies and organizational culture respect the right to self-determination and data sovereignty embodied by the Global Indigenous Data Alliance's CARE Principles (Collective benefit, Authority to control, Responsibility, and Ethics)[8], and specific manifestations, including the First Nations principles of OCAP® (Ownership, Control, Access, and Possession)[9]. The Alliance needs to support Indigenous communities in a manner consistent with the Truth and Reconciliation Commission's Principles of reconciliation[10] and the rights articulated in the United Nations Declaration on the Rights of Indigenous Peoples[11]. Digitization, preservation, and control over data have important complementary roles in cultural revitalization.

[7] Ontario Cybersecurity Higher Education Consortium (ON-CHEC). https://www.orion.on.ca/about-us/on-chec

[8] GIDA CARE Principles for Indigenous Data Governance. https://www.gida-global.org/care

[9] First Nations Principles of Ownership, Control, Access, and Possession. https://fnigc.ca/ocap-training

[10] Truth and Reconciliation Commission of Canada, Principles of Truth and Reconciliation. http://www.trc.ca/assets/pdf/Principles%20of%20Truth%20and%20Reconciliation.pdf

[11] United Nations Declaration on the Rights of Indigenous Peoples. https://www.un.org/development/desa/indigenouspeoples/declaration-on-the-rights-of-indigenous-peoples.html

Relationships and partnerships with First Nations, Métis, Inuit, and other Indigenous communities are essential to build a fully inclusive and equitable DRI in Canada. Importantly, the Researcher Council recognizes that it does not have representation of Indigenous researchers at this time and is actively working with the Alliance management to address this gap.

Priorities

- Provide funding support, expertise, and partnership to Indigenous Peoples in Canada so that they can develop their own sovereign digital research infrastructures and use digital methods in support of cultural revitalization.
- Facilitate opportunities for respectful collaboration which respond to and support community needs, goals, and aspirations.

4 Equity, diversity, and inclusion

The Alliance's mandate includes a commitment to integrate equity, diversity and inclusion (EDI) principles into every organizational initiative. *Equity* involves the fair and respectful treatment of all; *Diversity* requires drawing on the widest range of backgrounds, perspectives, and experiences, with attention to under-represented groups; and *Inclusion* enables full and respectful participation. A commitment to EDI involves providing support for researchers of all identities, cultures, and abilities, from across every research discipline, and in Canadian research institutions of all capacities from coast to coast. In the Canadian context, EDI must foreground the equitable treatment and inclusion of First Nations, Inuit, Métis, and members of other Indigenous communities.

The Alliance should acknowledge the barriers that many researchers experience in accessing and utilizing DRI. These include a lack of awareness of existing tools and services, little or no technical support at their institution, or being situated at a geographically remote institution. Systemic biases present in research and higher education institutions may also impact the access and use of DRI by marginalized communities. The Alliance needs to ensure that the tools and resources it provides respond to the needs of the broader research community.

Priorities

- Integrate EDI principles into all decisions and activities as a foundational principle.
- Adopt metrics for EDI that follow national and international standards and best practices, identify areas requiring improvement, address gaps promptly, and engage in continuous monitoring and improvement.
- Provide services and support to all researchers in French, English and other languages across Canada.
- Ensure all researchers, including those from marginalized groups, have access to the Alliance's tools and services, regardless of geography, discipline, or institutional affiliation.

5 Professional support personnel

Professional support personnel[12] are critical to the success of the Alliance. This includes data curators and data management specialists, system administrators, research software engineers and domain-specific support staff. The Alliance will inherit a strong team from Compute Canada. A comprehensive review of their areas of strengths and weaknesses will help the Alliance determine the optimal structure for providing support to the DRI ecosystem.

The need for local support was highlighted during the community consultations as paramount. Local DRI support personnel serve as a critical bridge with the national teams and provide the first point of contact for researchers.

In addition to the professional personnel supported by the Alliance, there is a vast pool of DRI expertise in the institutional IT groups, libraries, laboratories and research centres across Canada. Engaging and coordinating Alliance professional personnel with other teams is a long-term endeavour that will ultimately benefit the research community.

Current Compute Canada Federation staff are employees of the universities or consortia, and subject to the local employment agreements of the organization and provincial law. As the Alliance assumes responsibility for providing financial support for these professional staff, it can use its influence to ensure that personnel are treated professionally with competitive compensation and benefits, and offered job security in accordance with the human resource regulations of the external organization. Such efforts will help the organization and the Alliance retain staff in a very competitive market.

Priorities

- Ensure that there are sufficient professional personnel to provide the necessary foundational support to meet the diverse needs of the research community.
- Engage with the regional consortia, host sites, and other organizations to put in place a model for the ongoing support and retention of the professional support personnel, and establish a system for the training and education of students and early career staff.

6 Research data management and stewardship practices

Research data and its associated scholarly outputs are valuable research assets. With proper stewardship these assets have the potential to yield even greater impact through their sharing and reuse for the advancement of scholarship and innovation. The Alliance needs to build a national infrastructure to support standardized research data management and stewardship that

[12] Note that Statistics Canada defines HQP as being individuals with university degrees at the bachelors' level and above. https://www150.statcan.gc.ca/n1/pub/88-003-x/2007002/10331-eng.htm This is the definition used by the research community and the funding agencies. Hence, we do not use HQP in reference to the professional support personnel.

encourages common data models to foster national and international collaboration within disciplinary communities.

Data repositories, services, and platforms will need to enable researchers to adopt the FAIR Principles (Findable, Accessible, Interoperable, Reusable)13 and the Tri-Agency Research Data Management Policy14. The Alliance should ensure that the community has the tools and resources to meet the emerging requirements and associated ethical, legal, or commercial obligations for research data.

The stewardship of sensitive data requires special attention since it must be safeguarded against unauthorized collection, access, or disclosure. Health-related research often uses diverse data sources with varying degrees of sensitivity, such as clinical and administrative healthcare data, and population-level social and economic data. Support for the adoption of appropriate research ethics and privacy guidelines that adhere to regional legislation and practices will be required. The Alliance must engage with active stakeholders, such as governments and institutions, to address this issue.

Priorities

- Build infrastructure and support for the FAIR management, curation and preservation of research data in collaboration with Canadian universities and laboratories.
- Develop a strategy and plan for managing sensitive data, and collaborate with other national and regional groups active in this area, including Indigenous communities.

7 Education and training

Digital research infrastructure is complex and ever changing, and will require the ongoing training of researchers. There is a need to increase awareness among the research community of the benefits of DRI to research. Early career researchers, and those from academic disciplines that traditionally have not utilized DRI, may not know where to find the right solutions for their research. Increasing awareness will require efforts at the local, regional and national levels by the Alliance and its partners. Since researchers have different levels of technical knowledge, educational needs will range from introductory courses to specialized technical skill development opportunities. The Alliance and its partners will need to provide a wide range of education and training material such as web pages and manuals, training videos, courses, live chat for technical support, and summer schools. This will elevate research and innovation in Canada.

There will also be a need for training opportunities for the Alliance technical support personnel so that they can keep up-to-date in the latest technologies. The Alliance should foster an environment where learning and professional development are strongly encouraged.

[13] Wilkinson, M. et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018. https://doi.org/10.1038/sdata.2016.18

[14] Tri-Agency Research Data Management Policy. https://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

Priorities

- Develop a DRI training curriculum in partnership with Canadian universities that leverages existing programs, and is based on the priorities and expertise of the research community.
- Ensure that training programs and documentation are kept up-to-date, well-curated and easily accessible and are offered frequently.

8 Research software

Research software is an integral component of the software ecosystem required for academic research and DRI in general. Currently, Compute Canada supports the core software required for the fundamental elements such as operating systems and identity management systems. Compute Canada also supports the middleware layer that includes databases, data management and job management systems. The Alliance will need to continue the support of the core software and expand the middleware layer to utilize commercial cloud computing resources.

Research software is code and libraries written by or for researchers. Currently, CANARIE supports the development of research software and this responsibility will transition to the Alliance in 2022. Along with the core software and middleware, research software can be either open-source or proprietary. Open-source software is widely adopted and favoured by many research communities. Proprietary software is often expensive, yet some secure and professionally supported software is used widely in certain domains. The Alliance needs to consider how to support the development of research software, and establish policies regarding both open-source and proprietary research software developed in Canada and abroad.

Research software ranges from mature projects that support a broader community to the development of software that is exploring new avenues of research. Through partnerships, the Alliance should support and expand research software initiatives. CANARIE, for example, used its research software program to support existing projects or platforms, whereas researchers developing new and innovative projects find support from Tri-Council grants, CFI awards or from other organizations. The Alliance has a role in supporting the ongoing maintenance and upgrading of open-source middleware packages and mature research software. The Alliance's Inaugural Projects program can also help Canadian researchers develop software for the future needs (e.g. developing middleware to access commercial clouds or development of software for quantum computing or AI projects).

The Alliance needs to be aware that software, and especially research software, will continue to evolve rapidly with the underlying hardware technologies. Hence, it is critical that the Alliance build an agile and responsive team of professional personnel.

Priorities

- Establish support and ongoing maintenance of critical infrastructure software components.
- Review the research software needs of the community to determine the key open-source packages that need to be maintained by the Alliance's professional support personnel.

- Develop a strategy for supporting the development of new and innovative open-source research software.

9 National and international engagement

The Alliance needs to develop strong relationships with local, provincial and federal organizations, as well as representing Canada on the international stage.

The Alliance and CANARIE are funded together by ISED as part of the national DRI strategy. A strong relationship with CANARIE, which provides the national research and education network and links to international networks, will serve the Alliance well. As it expands its management of sensitive and private data, the Alliance will also need to leverage the growing expertise of CANARIE in cybersecurity.

The Alliance's contribution agreement with ISED stipulates that the federal government provides a fraction of the funding for Canadian DRI. As the Alliance builds essential partnerships with the provincial and regional DRI organizations, it is critical to remind them of the importance of researcher requirements and the priorities highlighted in this document.

The Alliance should continue to support specialized systems that are used to host unique data sets such as genomic libraries, or data from remote telescopes or accelerators. Humanities researchers have specialized needs for platforms (such as Voyant Tools[15] and the Canadian Writing Research Collaboratory[16]) that integrate multiple tools, coordinate collaborative activities and workflows, and support interactive and iterative research processes.

The Alliance should engage with organizations and initiatives that have DRI expertise in the health sector, such as the Health Data Research Network Canada[17], the Pan-Canadian Health Data Strategy[18], and the Canadian Institute for Health Information (CIHI)[19]. The Alliance may have the highest impact in this sector by focusing on specific technical elements of interest across Canada (such as linked secure health data repositories that are able to respect provincial and local boundaries).

Canadian researchers have collaborated with industry in many areas such as cloud and quantum computing, and data management systems. The Alliance should leverage the expertise of its research community, and its links to industry, with targeted projects in high-impact areas. Engagement with industry would be strongly supported by the federal government.

[15] Voyant Tools is a web-based reading and analysis environment for digital texts. https://voyant-tools.org

[16] Canadian Writing Research Collaboratory. https://cwrc.ca

[17] Health Data Research Network Canada. https://www.hdrn.ca

[18] Moving Forward on a Pan-Canadian Health Data Strategy. https://www.canada.ca/en/public-health/programs/pan-canadian-health-data-strategy.html

[19] Canadian Institute for Health Information (CIHI). https://www.cihi.ca/en

The Alliance can help Canadian researchers by learning how other countries and large international collaborations address their DRI challenges. To establish a global profile, the Alliance should participate in international DRI conferences and workshops, and visit Canadian researchers located at international research laboratories around the world. Building an international reputation will be a gradual process but it can ensure the long-term viability of the Alliance.

Priorities

- Establish relationships and partnerships with existing Canadian institutions, laboratories and industries to establish the Alliance's central role in the DRI ecosystem.
- Build connections with the international community that will help the Alliance develop new partnerships and support projects with a shared infrastructure.

10 About this document

The Researcher Council acknowledges the myriad of researchers in Canada who provided their perspectives, perceptions and time through the writing and submission of position papers, responses to a national survey and participation in a series of Town Halls. Foundational analysis of the input material used for this document was provided by the Alliance Senior Analysts. The first substantial draft of this document was written by Felipe Pérez-Jvostov, Shahira Khair (Alliance Senior Analysts) and Randall Sobie (Researcher Council Chair). Susan Brown, Laura Estill, Rebecca Davis and Amol Verma facilitated Researcher Council workshops and provided substantive contributions to the document. Editorial and critical review was provided by Karey Iron (General Secretary, Researcher Council), with contributions by Seppo Sahrakorpi and Qian Zhang (Alliance Senior Analysts). French context and translation review was performed by Researcher Council members Bruno Blais, Philippe Després, Benoît Dupont and Marie-Jean Meurs. External review was provided by Ry Moran, Associate University Librarian – Reconciliation, University of Victoria.

The Research Council is listed on the following page.

The Alliance Researcher Council (August 2021)

Karen Bakker
University of British Columbia

Rebecca Davis
University of Manitoba

Erika Merschrod
Memorial University

Girma Bitsuamlak
Western University

Philippe Després
Université Laval

Marie-Jean Meurs
Université du Québec à Montréal

Bruno Blais
Polytechnique Montréal

Erin Dickie
Centre for Addiction and Mental Health, University of Toronto

Erik Rosolowsky
University of Alberta

Guillaume Bourque
McGill University

Benoît Dupont
Université de Montréal

Rebecca Pillai-Riddell
York University

Susan Brown
University of Guelph

Laura Estill
St Francis Xavier University

Pekka Sinervo
University of Toronto

Carolyn Côté-Lussier
Institut National de la Recherche Scientifique

Benjamin Fung
McGill University

Randall Sobie
Institute of Particle Physics, University of Victoria

Constance Crompton
University of Ottawa

Anne Martel
Sunnybrook Research Institute

Amol Verma
St Michael's Hospital, University of Toronto

Carl D'Arcy
University of Saskatchewan