Open Scholarship mandates

As reflected in the publication of the Roadmap for Open Science, the sharing of products arising from publicly funded research require national structure and coordination. Open Scholarship mandates and policies should be developed in collaboration with national stakeholders to ensure alignment between compliance requirements and service provision across all levels.

While not all data can or should be openly shared (e.g., commercial potential or PII), appropriate mechanisms should be put in place to ensure this data can still be discoverable in a secure way, and in respect of Indigenous data governance.

3. Operations

Advanced Research Computing

The increased complexity and size of datasets and computational workflows has resulted in a growing demand for both CPU and GPU computing, large-scale parallel computation, and secure environments for the analysis of sensitive data. Oftentimes, this computing infrastructure also needs to be available for real-time processing, and in some cases even a 24/7 availability. Additionally, computing and storage requirements are projected to continue increasing well beyond current capacities. The Alliance should continue working with domain specific services and solution providers that focus on and provide large scale discipline specific systems and solutions.

For some disciplines, however, improvement in ARC does not necessarily involve increased CPU or GPU computing, but rather new and improved workflows and customized and integrated software. This especially the case in Digital Humanities, where workflows often include web hosting of specialized applications, software development, and data preservation services. Such workflows, platforms and applications development greatly depend on local Professional Support Staff and HQP support.

Cloud computing

Due to its decreasing cost and ease of use, there has been recently an increased adoption of commercial cloud services by academic researchers. This may suggest a gap in current service provision through national infrastructure, and thus a need to improve access to remote computing resources, low-cost data storage, curation, analysis, and burst computing. Support to accessing cloud-based solutions, whether through commercial or national platforms, should be done in tandem with better integration between institutional IT resources, and the national support teams. A more integrated DRI national ecosystem will make it easier for academic researchers to access the resources they require in a timely manner.

Collaboration with the Tri-Agency and Service Providers (national and regional)

Current discordance between Tri-Agency funding and availability of computational resources to carry out funded projects creates inefficiencies and delays progress. There is an important need to align with the Tri-Agency and ensure the availability of required DRI is synchronized with funding allocations.

Community engagement

The community's uncertainty about the Alliance and its future role in the national DRI ecosystem provides an opportunity to start anew the dialogue between service providers, funders, academic institutions, and the broader researcher community. Special emphasis should be placed on developing trustworthy partnerships with First Nations, Inuit, and Métis, and any other groups that have been historically under-represented. Recurring consultation cycles can facilitate users' satisfaction and trust, while ensure service provision responds to the community needs.

Cybersecurity

Increasing computer security personnel and cybersecurity education for researchers is essential to protect researchers' intellectual property. Researchers need more support for securely transferring and accessing sensitive data, as well as access to security resources, and information about best practices to secure research data.

Storage infrastructure is needed to handle sensitive data and allow researchers to enact the specific requirements of their research partners, communities, and institutions.

Equity, Diversity and Inclusion (EDI) across users of DRI, research disciplines, institutions, and geography

EDI is lagging across the many levels of the national DRI ecosystem. From low representation of historically excluded groups, women, and 2SLGBTQIA+, to the underutilization of the ARC infrastructure by many disciplines, attention should be placed in providing an equitable access to DRI, and the promotion of a more inclusive and diverse community of users.

International engagement

International collaboration is central to many disciplines. To be competitive at the global level, the national DRI should be capable of integration with international systems and enable international cooperation through the adoption of shared standards or policies for interoperability.

Long-term storage, data curation and preservation

Current funding streams rarely provide support beyond the active portion of a research project, which excludes the necessary support for long-term storage and preservation. This lack of funding opportunities to support core aspects of data curation is a consistent challenge across all research disciplines.

Figure 4. Question: Where is the computing infrastructure you use in your research? Check all that apply. *Multiple choice.* (Social Sciences and Humanities, n= 255; Health Research, n= 231; Science and Engineering, n= 620; Total = 1106)

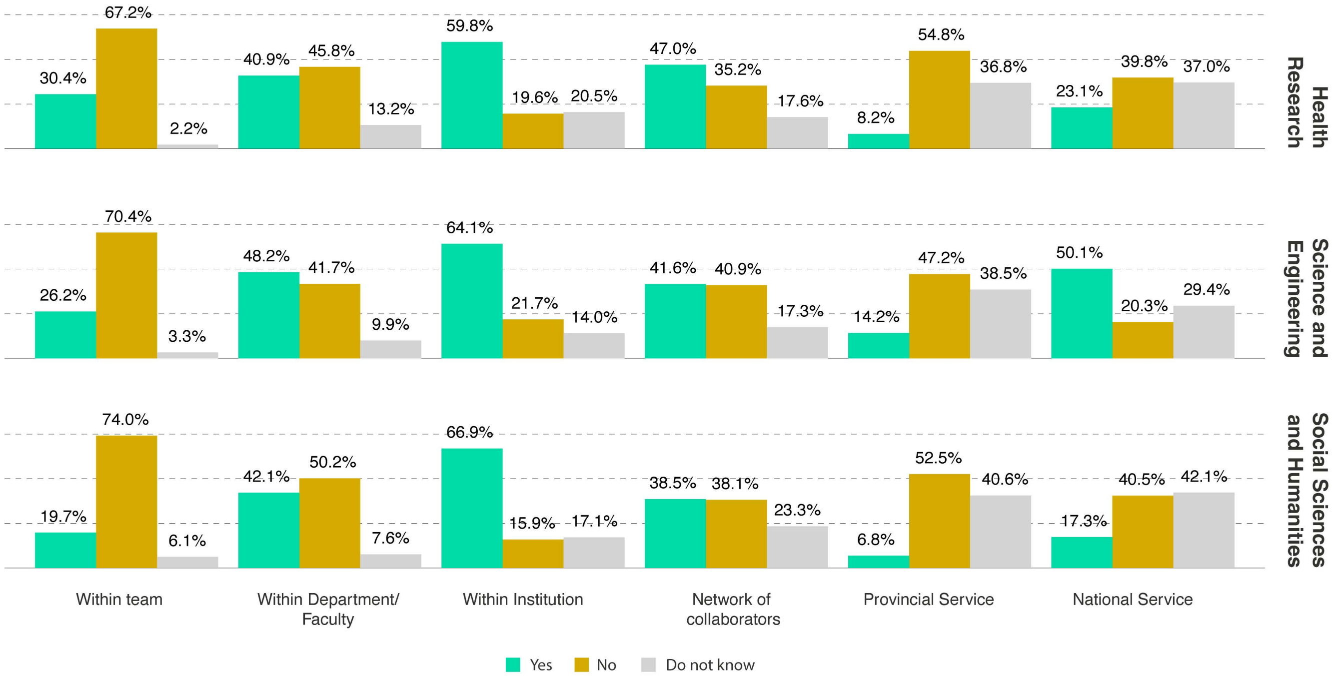

Figure 5. Question: Do you have access to professional IT staff to support your research workflows (e.g., Research Software Developers, System Administrators)? (Social Sciences and Humanities, n= 243; Health Research, n= 222; Science and Engineering, n= 592; Total = 1057)

When considering the means and mechanisms by which respondents discover and adopt tools and resources in their research, it was clear that peer-to-peer interactions are extremely important. For example, across disciplines peer/collaborators reference was the most important channel for the discovery and adoption of DRI tools across disciplines (Figure 6). This was also reflected in how researchers share and manage their data and code. Researchers from across disciplines identified that discipline adoption and collaborators' expectations are equally, or even more important, than funders or institutional requirements (Figure 7). Despite the importance of peer-to-peer interactions in driving community dynamics in relation to DRI, it is important to highlight that in the SE the most important channel for the discovery of tools and resources is through the CCF. Almost 66% of respondents from the SE discover tools through the CCF, compared to 33% in HR and 21.5% in the SSH (Figure 6). This lower adoption of tools through CCF in HR and the SSH is likely related to their lower usage of CCF platforms as previously described. Further efforts to increase access to and use of national platforms, such as the CCF, may also increase the adoption of the tools and services provided through them by these disciplines –a strategy that may also improve standardization and interoperability in these fields.

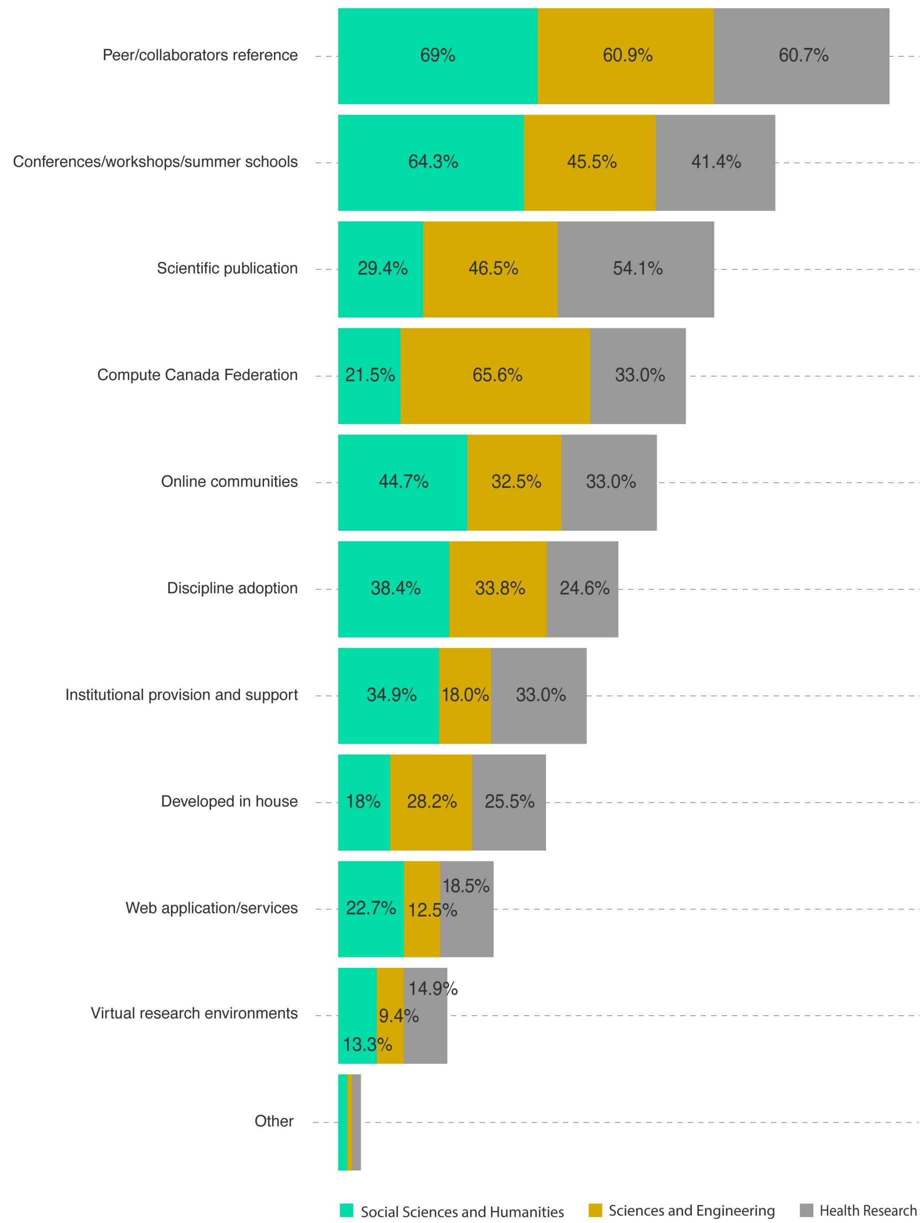

Figure 6. Question: Through which channels have you discovered the digital tools and services adopted or implemented in your research? Check all that apply. *Multiple choice.* (Social Sciences and Humanities, n= 255; Health Research, n= 227; Science and Engineering, n=615; Total = 1097)

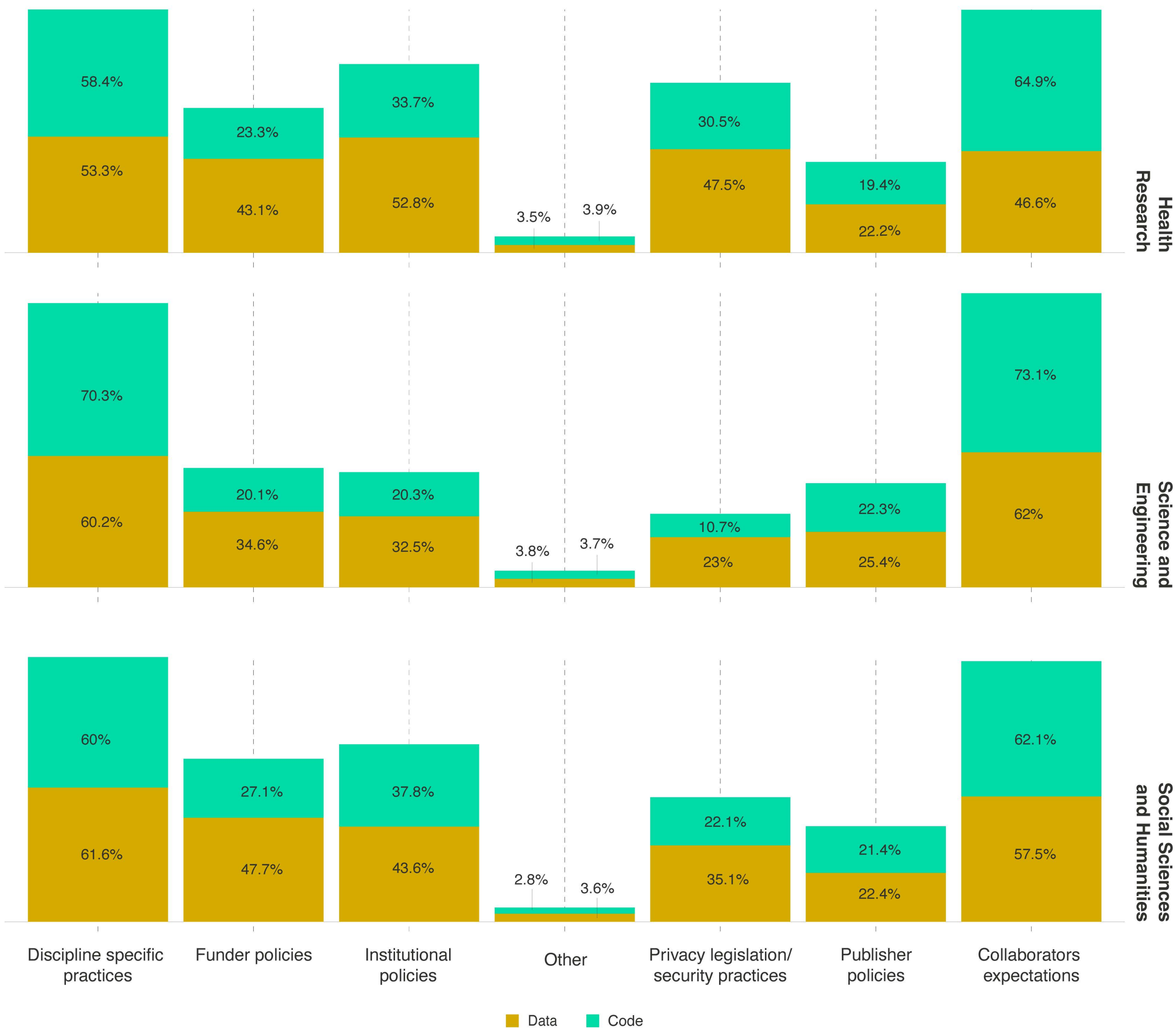

Figure 7. **Question:** Please indicate up to 3 items that influence your managing and sharing practices? (Social Sciences and Humanities, n= 245; Health Research, n= 225; Science and Engineering, n= 574; Total = 1044)

4 Support and Training

In addition to their usage of DRI, respondents were asked where they and their teams required training on the use of DRI tools and resources, and where they traditionally sought or received training from. 'The analysis of research data' was identified as the most needed form of training (Figure 8)–especially in the HR (47.8%) and SSH (62.5%). While training on the analysis of data was also important, other more *technical* training was prioritised by respondents in SE (e.g., using HPC systems, 49.6%; best practices on software development, 45.9%; parallel programming 36.7%). On the other hand, training on the archival and preservation of digital objects was an important training need in the HR (43.6%). Training in machine learning was identified as an important need in both SE (39.2%) and HR (43.3%). These training needs reflect the difference in the usage of DRI across disciplines, training offerings in undergraduate and graduate curricula, as well as the types of data and discipline/community culture.

Despite diverse needs in training between disciplines, researchers receive training from similar sources. Online courses, institutional training, and peer-led workshops all are important mechanisms by which researchers are trained in the DRI tools they use. (Figure 9). However, consistent with previous questions, over half of researchers in the SE identified receiving training from CCF, compared to only 10.9% in the SSH and 22.9% in the HR. Professional conferences were also important in the SSH (37.5%) compared to 19.2% and 26.1% in SE and HR, respectively.

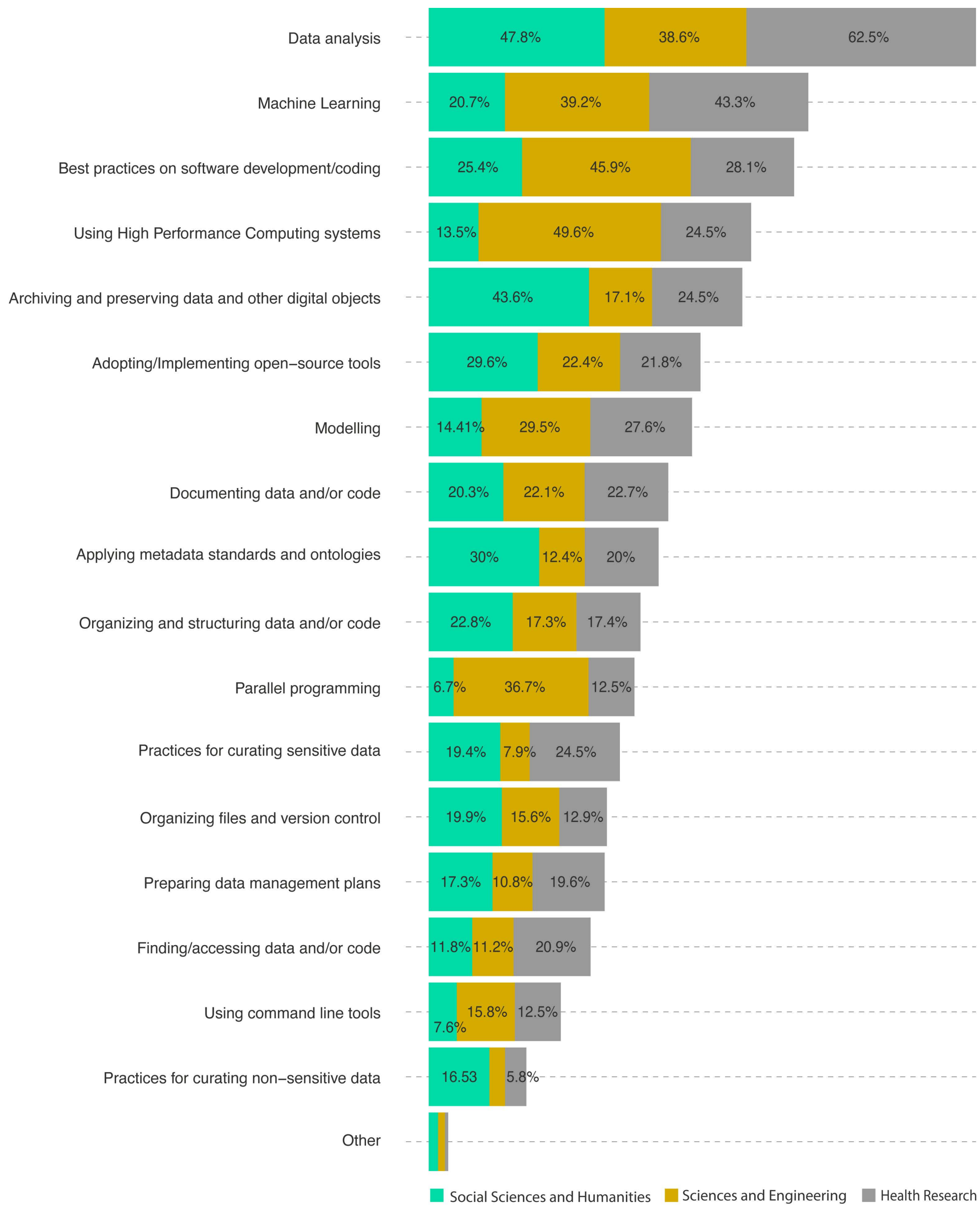

Figure 8. Question: Choose the top 5 areas where you, or members of your team, most require training to improve and maximize your use of DRI. (Social Sciences and Humanities, n= 236; Health Research, n= 224; Science and Engineering, n= 588; Total= 1048)

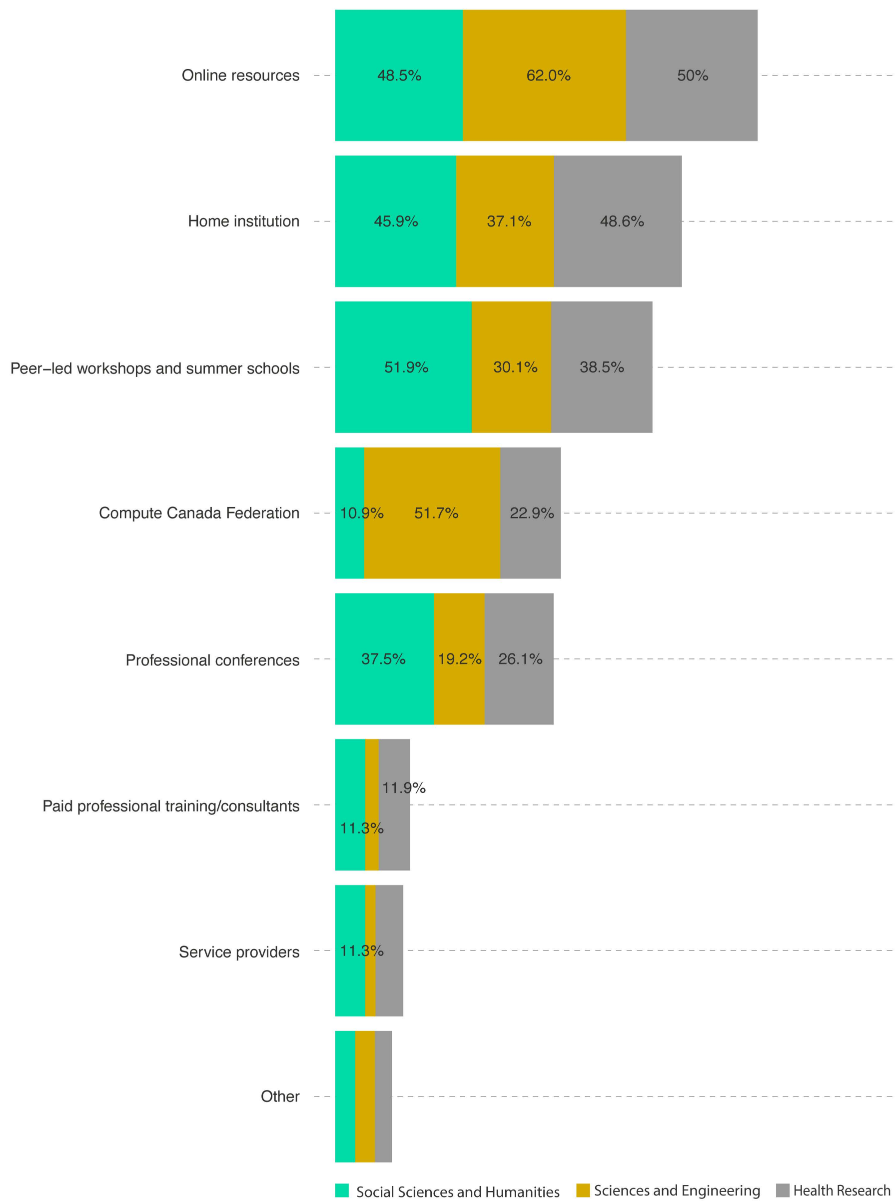

Figure 9. Question: Choose the 3 sources you, or members of your team, most often receive training from? (Social Sciences and Humanities, n= 237; Health Research, n= 218; Science and Engineering, n= 587; Total= 1042)

5 DRI needs and priorities

5.1 Needs

Researchers were also asked whether their DRI needs were currently being met, as well as what the Alliance should prioritize to fulfill their needs. The overall level of satisfaction with the current DRI was highest in the SE (Figure 10). Training and support were especially lacking in HR and SSH, where larger proportions of respondents either disagreed or strongly disagreed with these needs being met. This discrepancy between SE and both SSH and HR is consistent with the previous questions where most researchers in the SE reported receiving training and support provided by the CCF, compared to those in SSH or HR. Indeed, researchers in the SSH and HR mentioned that the main barriers they face in training and support are the lack of opportunities and training for highly qualified personnel (HQP). Poor documentation and too few workshops/courses were among the most common barriers. Many researchers from across disciplines also described important barriers at the institutional level, where administrative barriers and inadequate support for the use of DRI limited their ability to access and utilize DRI. Many more were not aware of the tools and services available to them.

A consistent low-level of satisfaction with current funding for DRI was identified across disciplines. Researchers commented that despite the significant costs associated with maintaining professional staff to support their DRI projects, many DRI-associated costs are not covered through their traditional research grants. Many other researchers also noted that their home institutions do not offer funding for DRI initiatives, and that personnel and infrastructure funding is reliant on temporary grants which are extremely competitive. Current salaries and funding options to attract and retain HQP and Professional Support Staff were also identified as not being competitive enough with those of the private sector. These positions were also mentioned as term positions with little to no job security.

Roughly a third of researchers across disciplines highlighted that their computing needs are not currently met. Many researchers alluded that CCF's Resource Allocation Competitions (RAC), which are based on project cycles, are cumbersome and do not promote building shared infrastructure and long-term scalability. Many others, in the SSH and HR in particular, discussed the vast learning curve that exists, and the amount of knowledge required to properly engage with DRI and technology as being a barrier to accessing systems. Many researchers described accessing CCF infrastructure as difficult, riddled with long wait times and unreliable access. Many others indicated that they are not aware of what specific DRI resources and support systems are available to researchers in Canada.

5.2 Priorities

While there are important differences in the usage of DRI, and hence the DRI-related needs across disciplines, funding for professional research support staff (SSH= 46.2%; SE= 30.5%; HR=

43.9%) and training for HQP (SSH=43%; SE= 37.4%; HR= 37.1%) were identified as priorities across disciplines. Other priorities across disciplines included cloud computing (SSH=32.6%; SE= 32.5%; HR=40.7%) and repository storage for data sharing (SSH=37.2%; SE= 28.9%; HR= 38.3%).

Differences in priorities were also observed. For example, researchers in the SE prioritized computational resources (e.g., high-speed internal networking computing= 48.7%, standard computing= 34%); while researchers in the SSH prioritized funding for online knowledge mobilization (43%) and archival storage (37.6%). Researchers in HR further identified secure storage for data as a priority (35.8%).

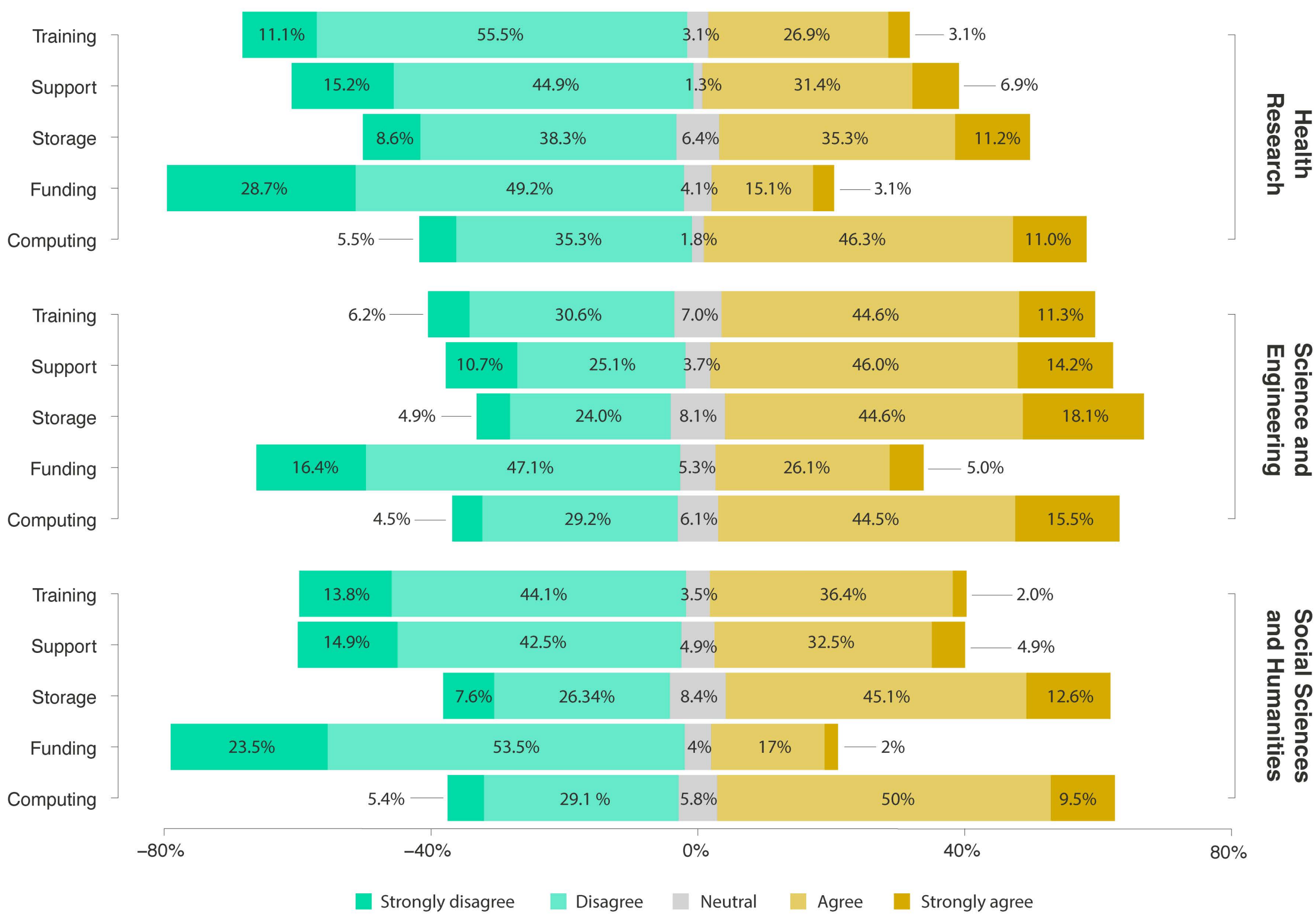

Figure 10. Question: How well are your current Digital Research Infrastructure needs being met? Semantic differential. (Social Sciences and Humanities, n= 224; Health Research, n= 210; Science and Engineering, n= 535; Total= 969)

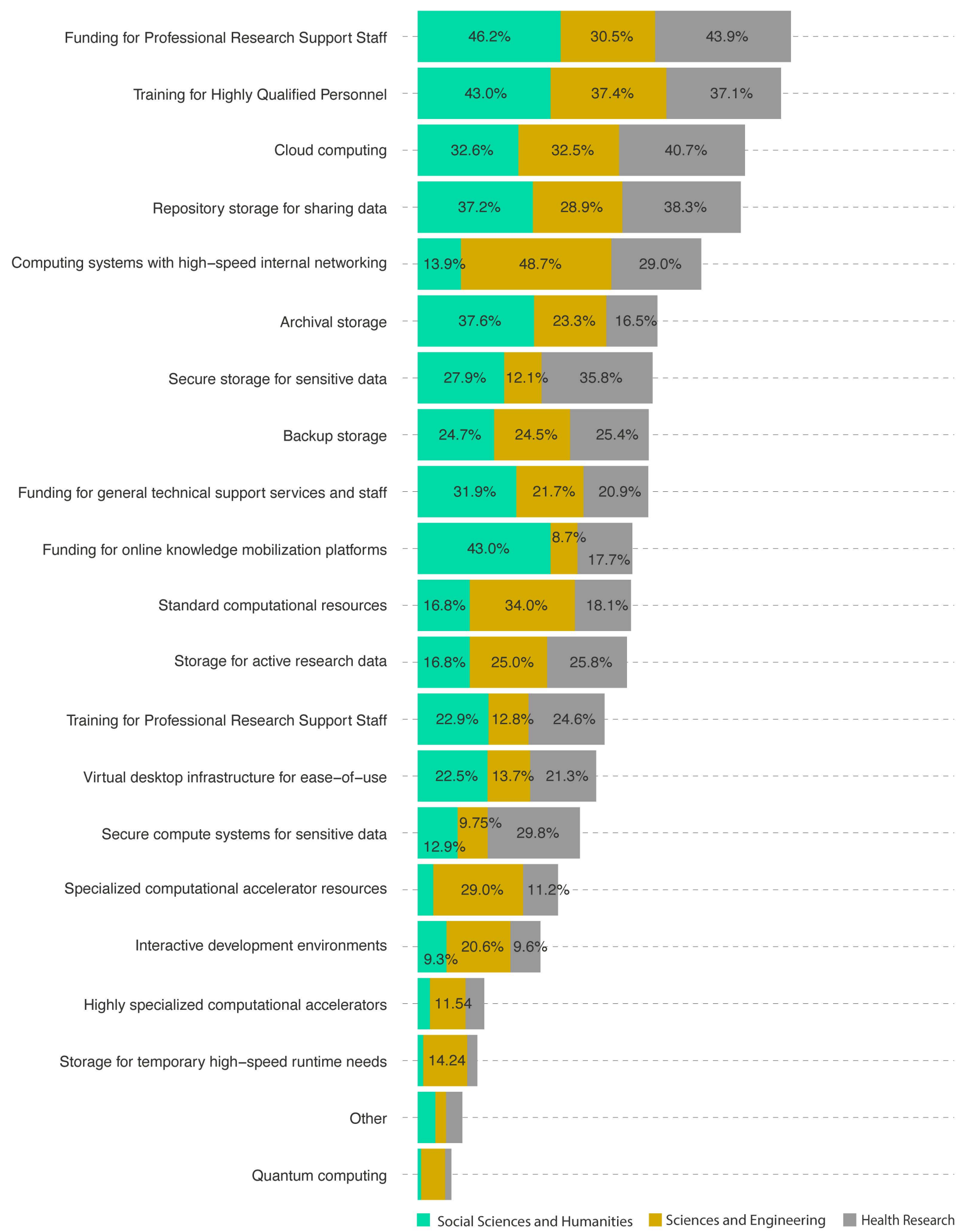

Figure 11. Question: What should Alliance prioritize to meet your current and future DRI needs? Choose the top 5. *Multiple choice*. (Social Sciences and Humanities, n= 279; Health Research, n=248; Science and Engineering, n= 667; Total= 1194)

6 Recommendations

Respondents were also asked to describe in free open text form how the Alliance could improve the current DRI in Canada. The following themes were identified (listed below in alphabetical order):

- Adopt and promote the use of free and open-source hardware and software.
- Be transparent.
- Coordinate DRI funding opportunities with the Tri-Council.
- Establish a funding program dedicated to cover salaries for Highly Qualified Personnel (HQP) and Professional Support Staff.
- Expand on discipline-specific technical support.
- Expand the provision of training on DRI tools and resources, from basic to advanced.
- Improve outreach and engagement of under-represented/under-utilizing communities and bring them to par with traditional DRI users.
- Improve remote access to DRI.
- Improve support for small Advanced Research Computing (ARC) users.
- Increase support in the Social Sciences and Humanities (SSH) and Health Research (HR).
- Improve the usability of DRI (e.g., ease of use)
- Make DRI free and accessible to every researcher in Canada, regardless of institution, geographic region, or research discipline.
- Provide professional services for the development of code, algorithms, and pipelines.
- Remove administrative barriers to access and use of DRI.

The Alliance Survey participation by region

Institution	n
Alberta	
University of Alberta	47
University of Calgary	21
University of Lethbridge	14
Athabasca University	7
Concordia University of Edmonton	2
Lakeland College	1
MacEwan University	1
NorQuest College	1
British Columbia	
University of British Columbia	88
University of Victoria	48
Simon Fraser University	18
TRIUMF	9
Thompson Rivers University	3
University of Northern British Columbia	2
University of the Fraser Valley	2
Population Data BC	1
Royal Roads University	1
Manitoba	
University of Manitoba	32
University of Winnipeg	10
Brandon University	4
New Brunswick	
Université de Moncton	7
University of New Brunswick	5
New Brunswick Community College	2
Mount Allison University	1
Newfoundland and Labrador	
Memorial University of Newfoundland	24
Nova Scotia	

Institution	n
Dalhousie University	24
Saint Francis Xavier University	7
Saint Mary's University	6
Acadia University	3
Cape Breton University	2
Nova Scotia Health	1
Ontario	
University of Toronto	185
York University	75
Western University	57
University of Waterloo	42
McMaster University	41
University of Guelph	38
University of Windsor	24
Queen's University	19
University of Ottawa	16
Ryerson University	15
Lakehead University	11
Carleton University	7
Ontario Tech University	6
Perimeter Institute for Theoretical Physics	6
Wilfrid Laurier University	6
George Brown College	5
Laurentian University	5
Brock University	4
Trent University	3
Ontario College of Art & Design University	2
Niagara College	1
Northeastern University	1
Ontario Health	1
Princess Margaret Cancer Centre	1
Unity Health Toronto	1
Women's College Hospital	1
Prince Edward Island	

Institution	n
University of Prince Edward Island	7
Québec	
McGill University	127
Université de Montréal	70
Université Laval	38
Concordia University	35
Université du Québec à Montréal	27
Université de Sherbrooke	11
Polytechnique Montréal	9
Université du Québec à Rimouski	9
École de Technologie Supérieure	7
Université du Québec à Chicoutimi	7
Université du Québec à Trois-Rivières	6
Université du Québec en Abitibi-Témiscamingue	5
Institut national de la recherche scientifique	3
Ouranos	2
CÉGEP de Sherbrooke	1
CÉGEP de Trois-Rivières /CMQ	1
Centre de recherche informatique de Montréal	1
Collège de Bois-de-Boulogne	1
Institut de recherche Robert-Sauvé en santé et en sécurité du travail	1
John Abbott College	1
Université du Québec en Outaouais	1
Université TÉLUQ	1
Saskatchewan	
University of Saskatchewan	21
University of Regina	10
Saskatchewan Polytechnic	1

Table 1. Respondents by region and institution.

Appendix 2

The Alliance's Needs Assessment Position Papers Report

Table of Contents

1 Introduction and background

The first consultation mechanism of the Canadian Digital Research Infrastructure Needs Assessment was a call for position papers from the general Canadian research community and was especially envisioned as a clean slate upon which authors could share with the Alliance their perspectives, needs, and recommendations for a cohesive and equitable DRI landscape. The call was open between October 22nd and December 14th, 2020. A total of 107 unique papers were submitted by 350 authors from across 112 organizations. The full list of submissions is presented at the end of this appendix, with direct links to the original document (see Position Papers).

All position papers were collated and categorized based on topic and general scope. Once categorized, the main takeaways and recommendations were extracted and organized by the Alliance's RDM and ARC working groups. Findings were then re-organized by theme and topic, and presented here as topic-specific needs. Recommendations submitted in the Position Papers are presented as community recommendations in the relevant topic-specific need. References to specific papers are shown within parenthesis.

2 Awareness and Accessibility

2.1 Awareness of existing infrastructures

Several submissions highlighted that many of the tools, services, and resources that researchers need already exist (01, 13, 15, 44, 02, 62, 76), but awareness of where to find them and how to access them varies across different communities, and is especially low outside Science, Technology, Engineering, and Mathematics (16). Even for researchers well versed with DRI, in the current Canadian context, it is quite challenging to navigate the DRI ecosystem. "*The main challenge for any given researcher is finding the pieces of the puzzle that they need and putting them together to form the picture that they require to accomplish their research*" (13).

Some authors mentioned that accessing Compute Canada Federation's infrastructure can be complicated compared to setting up commercial cloud services, especially for short-term needs (2; see Commercial Cloud).

Community recommendations:

- Collaborate with entities at the institutional level to help raise awareness about national DRI resources (20, 54, 02, 103). For instance, the submission by Ryerson University's Centre for Digital Humanities presents a vision that sees the Alliance as mediated through a constellation of university research centres and libraries (91). Institutional ambassadors paired with the Alliance's staff could lead outreach (54).

- Offer institution-specific portals to support multiple needs, including awareness, accessibility, and/or monitoring (76, 56, 41).
- Develop and maintain an inventory/catalogue of existing infrastructures (01, 13, 15, 44, 02, 62, 76).
- Athabasca U. describes the range of tools, national tools (e.g., Compute Canada, WestGrid), provincial tools (e.g., Cybera), and institutional tools, and their recent shift to Amazon's Web Services cloud for operational computing. Decision making tools to help identify which resource is best suited for a project would be helpful to prepare project proposals and resource plans (40).
- Improve access to commercial cloud by negotiating bulk prices and act as on-ramp for commercial services (2).

2.2 Inequality in access to institutional support

While many groups may not be aware of national DRI resources, like Compute Canada or Portage, they do often seek support from their own institutions, including research centres and libraries (see Institutional support). Several authors highlight important differences in the level of support available across institutions, which reduce access and use of available tools and resources, and even leads researchers to arrange their own DRI. This has important implications for the sustainability and persistence of data and infrastructures (16, 18, 27). For instance, the burden of supporting Research Data Management at the institutional level that is currently being placed on libraries can augment inequities between institutions, especially on smaller institutions and colleges that may lack the personnel capacity to support their researchers. Authors from the humanities also note turning to the libraries for support with complex digital projects, and while it may not be part of the traditional mandate of libraries, this support is nonetheless essential to the digital humanities.

Community recommendations:

- Access to resources should be equitable regardless of institutional size or location (22, 53, 56, 101). An example raised by UQAC was that their researchers need to travel to larger urban centres to access certain datasets provided by Canadian Research Data Centre Network's (CRDCN) research data centres (RDCs) for accessing highly sensitive Statistics Canada data. This is an undesirable situation to repeat for the Alliance's offerings.
- Representation of smaller institutions in the Alliance's decision-making committees is an important consideration for equity and inclusion (56).

2.3 Inequality in access to funding

Three main areas of inaccessibility to funding were identified in the submissions. These were related to discipline, institution, and community. Confusion regarding differences in rules for matching funds between provinces and research disciplines were highlighted as direct drivers of inequities (55). For example, submissions from researchers in the Social Sciences and Humanities (SSH) noted that grants awarded by Social Sciences and Humanities Research Council (SSHRC) explicitly disallow funding infrastructure and have significantly lower caps on software development allocations than the other Tri-Agencies, impacting SSH researchers disproportionately (18, 55). On the other hand, Natural Sciences and Engineering Research Council (NSERC), which allows funding for infrastructure, will not fund humanities research (18). Others noted that the complexity of Canadian Foundation for Innovation (CFI) applications and resulting high failure rate suggest this application process is out of reach for many researchers (102).

At the institutional level, submissions from smaller universities reiterated the importance of national services, since these offer the support that they are unable to fund locally and require equitable fee models to access them (76, 101).

The First Nations Information Governance Centre (FNIGC) addressed the need for access to equitable funding for First Nations to collaborate with the Canadian data regime while developing their own data structures that can interface with the Canadian system and protect First Nations data on their own terms (98). A paper on reconciliation by Ry Moran also noted the need for targeted funding opportunities for Indigenous groups (102).

Community recommendations:

- Following publication of the Truth and Reconciliation Commission Report, the Tri-Council established targeted funding to support research on reconciliation and to support Indigenous research and research training. More is needed to identify and address community and intergenerational harms in research, including directed funding to communities to support their research goals and data sovereignty (39).
- Consolidate equitable and agile funding streams that respond to the needs of researchers from across disciplines (55).

2.4 Bilingualism

The most frequently raised point regarding equity and inclusion concerned access to French-language services and support. Several submissions reiterated that French should not be an after-thought but an integral part of infrastructure and project planning from the start (08, 49, 56, 58, 64, 65, 82, 87, 89).

Community recommendations:

- Internalize the need for bilingualism to penetrate to tools and services, with respect to interface, documentation (on guides, modules, policies, etc.), training and support.
- A submission from UQAC suggests a formal partnership between the Alliance and Calcul Québec to support its French-language services (56).

2.5 Accessibility and usability of infrastructure

Many authors emphasized the importance of easily accessible tools, especially for storage, collaboration, sharing and preservation on trusted and sustainable platforms to create a stable foundation for research in Canada. This democratization of access to the data and to computing and analytical resources would contribute to creating equity–Canadian researchers need to be able to compete not only in the international scene, but also with private, multinational giants that have access to nearly unlimited compute power (84).

Recommendations:

- A single interface should be provided to link to as many Alliance's administrative functions as possible (54).
- Provide basic collaborative platforms with low barriers to use, both in terms of usability and cost, are needed to support wider accessibility to DRI (10, 15, 20, 54).
- *"Interfaces through which users interact with Alliance's technologies will shape the future of Canada's research community."* The Report from Antoniuk and Brown argue that while command-line tools dominate many traditional DRI contexts, thoughtfully designed graphical user interfaces (GUIs) can support greater accessibility and equity for tools and services (54).
- Build on Compute Canada Federation (CCF) achievements, such as the Research Support and other National Teams, to provide more uniform service across sites through the helpdesk, wiki, software environment–even though there are still many areas of differences between sites that pose barriers for access and usability (93; see Highly Qualified Personnel).

3 Governance and Policy

3.1 Accountability

Several submissions discussed the role of the Alliance as an emerging organization and the values it ought to adopt. While accountability for the use of public funds was described as essential for a national organization as the Alliance, it should be done with as little overhead as possible. The administrative overhead associated with some components of the current funding model are so high that many researchers and organizations would be deterred from participating in the Alliance's initiatives, resulting in reduced innovation, synergy, and alignment of efforts (54).

Community recommendations:

- Given the amount of power the Alliance will exercise with respect to data infrastructure, data management, re-use, and related innovation, there should be structures and processes to ensure full accountability to the public that funds the Alliance, and the majority of its member organizations (72).
- There is a danger of corporate influence on and control of research, and on decision-making about research infrastructure (*see Private Sector Influences below*, 72). Therefore, the Alliance should develop clear accountability mechanisms that involve the public in governance of DRI and in research and procurement decisions (72).
- Key Performance Indicators (KPIs) should be developed for operations and made publicly accessible (72, 92).
- Given the current climate crisis, the Alliance should be transparent about the greenhouse gas emissions of ARC facilities and publicize expectations around net-zero targets (19).

3.2 Collaboration and trusted relationships with Indigenous Peoples

The FNIGC highlighted that a national strategy for DRI must take into consideration the policies, processes, and people necessary to establish and maintain trusted relationships with First Nations researchers and communities, and that collaboration and co-development must be prioritized (98). Similar efforts ought to be made to develop relationships with Inuit and Métis researchers and nations. Promotion and inclusion on decision-making bodies is important for relationship building (e.g., the Alliance Researcher Council), and it would allow for the representation of different levels of experience, not only in technical expertise, but also real-life experience of how Indigenous Peoples view data, and the unique ways Indigenous data and people are at risk within the current Canadian DRI ecosystem. Without this experience, it is impossible to recognize implicit biases policies and systems may present (98).

Community recommendations:

- New models for collaborative co-stewardship of research data need to be developed through reciprocal partnerships between Indigenous Peoples and research organizations, which are situated in local relationships and ways of knowing (36).
- There is an important role for the Alliance to play in advancing reconciliation. The paper submitted by Ry Moran asserts "Canada's national digital research infrastructure must at a minimum embrace a complete and comprehensive project of preservation and revitalization that supports the resurgence of Indigenous cultures" (78).

- Promote the inclusion of First Nations, Inuit, Métis, and other Indigenous Peoples in Canada in the development of the national DRI strategy (98).

3.2.1 Indigenous data sovereignty

Indigenous Peoples need to be engaged in an active dialogue on how to protect their own data, while allowing for interaction of that data within the Canadian DRI. The Alliance has a leadership role to play in advancing priorities of Indigenous Peoples through endorsement of the Collective Benefit, Authority to Control, Responsibility, and Ethics (CARE) principles and establishment of Indigenous-led DRI that leads to the development of community capacity and advance of data sovereignty and unique governance mechanisms (01, 35).

Successful examples of collaboration and co-development include the work of ICES (formerly known as Institute for Clinical Evaluative Sciences), which has established an Indigenous Portfolio that works closely with First Nations, Inuit, and Métis, organizations in Ontario in partnerships that include data governance and data sharing agreements, which has enabled Indigenous-driven research and data analysis using data held at ICES.

Community recommendations:

- The Alliance's policies, support and tools need to be co-developed so that data sovereignty needs are taken into consideration from the start (36, 39, 94, 98, 102).
- Create opportunities for First Nations to appoint technical advisors and political leaders to represent interests of their respective communities in dialogue with Canadian experts and data leaders.
    - *Example:* joint technical table of First Nations could address the two key issues of:
        1. development and implementation of Canadian systems, tools, and protocols that respect First Nations data sovereignty, and
        2. development and implementation of First Nations systems, tools, and protocols necessary to exercise their data sovereignty (98).
- Create an Indigenous Advisory Council to guide the Alliance on digital research infrastructure policies involving the preservation and protection of Indigenous data and its sovereignty (36).
- The Health Data Research Network envisions a future state where data related to First Nations, Métis, and Inuit populations has established Indigenous governance and control by Indigenous organizations to inform decision-making (35).

3.3 Sharing PII and health data

Several authors highlighted the difficulties that researchers face when searching for storage and computing solutions for sensitive data (e.g., personally identifiable information (PII); health records), among which were the strict requirements for compliance with related provincial and national legislation (32, 86). Below are the main points related to legislative and regulatory barriers.

3.3.1 Legislative barriers

Current architectures and regulations limit Canadian researchers' participation in the global shift towards sharing and linking large-scale health, clinical, and genomics data (61, 86). Specifically, provincial and international differences in legislation are important barriers for the access to, as well as the linking and sharing of research, clinical, and administrative data (05, 11, 32, 35, 61, 85, 86, 96). The submission from Sick Kids additionally noted an absence of reciprocal agreements between research sites that address regional differences in regulatory frameworks (e.g., The Personal Information Protection and Electronic Documents Act (PIPEDA), federal; The Personal Health Information Protection Arc (PHIPA), Ontario) for clinical data management (86). Access to infrastructure that is compliant with the different laws is a key consideration for collaboration with other sites, as noted in many other submissions.

Community recommendations:

- To facilitate health research in Canada, the Alliance should support secure computing and storage architecture that is compliant with relevant privacy legislation and directly available to Canadian researchers (15).
    - This could be partially resolved by greater support for provincial or local storage options (101).
- Authors from a report by the Canadian Health Workforce Network suggest "*a commissioned legal review and set of suggested amendments to legislation, regulation and organizational by-laws, paralleling existing reviews with regard to patient data*" (11).
- Health Data Research Network Canada (HDRN) is undertaking a review to distinguish between legislative requirements vs. regulatory practices vs. procedural choices of jurisdictions and lay the foundation for future regulatory reform (35).
- International legislative requirements are another consideration in the collaboration and use of infrastructures. In particular, the implications of the General Data Protection Regulation (GDPR) and GDPR-like models should be considered with stakeholders in national discussions/working groups (35, 47).

- Building upon The Global Alliance for Genomics and Health (GA4GH) standards will enable both national and international collaborations (12).

3.3.2 Regulatory barriers

Administrative regulatory policies were identified as another important challenge for health researchers, especially because they limit extraction of value from the data that can be shared (61). A report from Diabetes Action Canada noted that each provincial organization holding and managing health data has their own policies and procedures for data quality and standardization. Some authors further considered that data sharing between researchers, either within or between individual institutions, is complicated using separate computational and storage systems, and the need for both inter-institutional data use agreements and forward-looking research ethics approvals that explicitly contemplate broad data sharing (86). Data trusts were discussed in several submissions as alternative structures to current data access models (04, 07, 35, 61, 85; see Data trust environments).

A more fundamental challenge noted by the Canadian Research Data Centre Network (CRDCN) are regulations that prevent government departments from making administrative records available for analysis (96). Authors from western health research organizations emphasized that these challenges arise due to cumbersome or outdated governance structures, which are intended to ensure confidentiality and conformity to privacy requirements, including ethics boards and data access committees, but which prevent timely access to data (04, 35). This has important implications for the management of incidental findings in the absence of informed consent, which is regulated by the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS2) and Research Ethics Board (REB) review process, and that prevents contacting patients and health care providers with potentially life-saving information (05, 61).

Specific examples were also provided. For genomics and health data, the highlight was on the need for a scalable, pan-Canadian federated approach that enables the query and analyses of national-scale genomics and human health datasets, while the data itself remains securely and privately controlled within research hospitals and other organizations across the country (12). Specifically, the authors mention that polices, governance and architecture, application programming interfaces (APIs) and ontologies tie all participating sites together into a single coherent network: HPC4Health (Ontario) and SD4Health (Québec) are two health computing nodes already collaborating with one another and with the BC Cancer Genome Sciences Centre to develop common APIs for data access through CanDIG (see Management of sensitive data).

Community recommendations:

- Secure repositories are needed to facilitate data sharing or data visitation over the long-term through controlled access to systems that are trusted and approved by relevant "data custodians" (61).
- Coordination is needed around data integrity and security protocols to satisfy data sharing agreements and appropriate research ethics guidelines (86).
- Overarching governance and guiding principles for institutional research ethics boards for data and image sharing are needed (84).
- Informed consent language and processes should be harmonized across organizations to support data sharing, as appropriate (12, 52, 39).
    - Standardized medical consent ontologies would also prevent regulatory violations (12).
    - More practical mechanisms to obtain consent for use from patients should be developed.

- For example: *"(digital) consent, multilingual and multicultural support and dynamic consent practices that allow patients to control how their data is used"* (05).

3.3.3 Race-based data

The risks to patients and communities presented by race-based data collected and sold in the process of research are summarized in a submission by LLana James (79).

Community recommendation:

- The Alliance should refrain from supporting the collection, linkage, or transfer of race-based data within its DRI until updated regulations and laws are in place to address specific risks at the intersection of race-based data (79).

3.3.4 Alternative governance models

Given that data governance goes beyond access, and must also address data acquisition and preparation, stakeholder and public engagement, and capacity building, several authors highlighted the need for infrastructures where protocols for storage and access can be co-created and negotiated with data owners (e.g., data from Indigenous Peoples, personal health data; 81, 85, 94, 98, 102). Agreements and support for improved data governance models were similarly mentioned.

Community recommendations:

- A range of models are recommended for data repositories to support different communities, ranging from models where access is provided on case-by-case by data stewards, to models involving rapid automated decision making (04).
- Adopt the Five Safes framework for HPC in relation to data privacy and protection requirement across provinces/territories: Safe Projects, Safe People, Safe Settings, Safe Data, Safe Outputs (07, 35).

3.4 Alignment with Funder and Institutional Policies

Areas around the Canadian Common CV, RDM Policy, and high-performance computing funding were often cited as important areas for improvement (08, 37, 39, 41, 54, 58, 62, 64, 92, 103).

CCF's data storage policy that requires yearly renewal of storage space conflicts with existing retention requirements from the Tri-Agency (34) (e.g., Canadian Institutes of Health Research

(CIHR) Open Access Policy on Publications, SSHRC Data Archiving Policy). The recent Tri-Agency RDM Policy will also present new implications around the data deposit requirement.[3]

Community recommendations:

- Partner and closely collaborate with the national research funders as means to harmonize practices and policies (08, 37, 39, 41, 54, 58, 62, 64, 92, 103).

3.5 Collaboration with industry

The importance for national digital research infrastructure supporting data sharing among the public and private sectors was identified as paramount to convert investments in innovation at universities into economic benefits for Canada (12, 61). Nonetheless, despite commercial data being a common data source in many fields of research, several authors highlighted that such data often cannot be openly shared (37, 77, 96). Others highlighted that commercial publishers who have traditionally acted as gatekeepers to publicly funded research, have now increasingly advanced their repository and RDM-support offerings (47).

3.5.1 Computing for Industry-Academic Partnerships

Opportunities for access to computing for industry-academic projects are not open enough to support benefits of research to industry, business, and society (77). These may include forming relationships with university-based accelerator programs (85). For example, the submission from the Southern Ontario Smart Computing Innovation Platform (SOSCIP) Consortium explains that since priorities of companies are apart from the research ecosystem, these systems should be separate from those dedicated to “pure discovery-based” research. Nonetheless, dedicated resources are needed to support collaborations “at the speed of business” (77).

Community recommendations:

- DRI services, support, and resources must be planned with industry/academic partnerships in mind.
- DRI investments should include support for industry-led ARC infrastructure. Dedicated platforms to support industry with high level security are needed, and software and services must be provided to help migrate to commercial platforms after projects are concluded, otherwise project may never get underway since the results are stranded on the system and would not be able to commercially benefit the company (77).

---

[3] Government of Canada: Tri-Agency Research Data Management Policy https://science.gc.ca/eic/site/063.nsf/eng/h_97610.html (March 2021).

3.5.2 Private sector influences

A submission from a diverse group of researchers in Ontario raised concerns related to corporate influences on research and on decision-making about research and research infrastructure (72). The negative impact of these influences was exemplified by challenges in accessing and sharing data from medical devices and primary care Electronic Medical Record (EMR) and data from private sector EMR vendors (05, 07, 52). In particular, authors from Diabetes Action Canada (5) explain "*EMR vendors have practices that limit the access to patient data for research, through additional costs, tolls and delays. It is a pervasive practice in the industry to impose fees for access to a clinic's or patient's own data for secondary use, including research, epidemiology or health system monitoring*". These practices are exacerbated by the absence of legislation mandating interoperability, preventing private sector from blocking data sharing, monetizing access to health data, and the lack of oversight of private sector activities (7, 47, 52)

Community recommendations:

- Establish structures and processes to ensure full accountability to the public and its member organizations, "*given the amount of power the Alliance will exercise with respect to data infrastructure, data management, re-use; and related innovation*" (72).
- Access to data arising from non-commercial sources should not be made available to representatives of private/commercial enterprises under the same conditions of access as to independent researchers (72).

3.6 Open scholarship mandates

The recent United Nations Educational, Scientific, and Cultural Organization (UNESCO) draft Recommendation on Open Science advocates for member states to "actively engage in removing the barriers and disincentives for Open Science, particularly those relating to research and career evaluation and awards systems." Many authors argued that such open access to scholarship, including research data, should be prioritized in policies, infrastructure, and services provided by the Alliance (e.g., 08, 10, 33, 66). However, many noted that a universal adoption of open scholarship without consideration of who decides what is shared directly conflicts with the data sovereignty of Indigenous Peoples (98), data privacy, and intellectual property. For example, qualitative researchers may have specific concerns and hesitation about open science/data sharing (94). Data ownership policies (affecting researchers and subjects/participants) require careful consideration, with implications on who has the right to grant access to the data in the future and what controls are required (94).

Several submissions identified the Tri-Agency Policy on Research Data Management as a tipping point that will spur institutional strategies and policies. The Alliance would thus be in position to lead on accountability with these open science mandates by developing mechanisms to monitor and incentivise institutional and researcher compliance (15, 41, 62).

Community recommendations:

- Consider open scholarship mandates and policies of ecosystem partners. For instance, the paper from the Dynamic Brain Circuits group from University of British Columbia (UBC) notes that the Compute Canada Federation's data storage policy that requires yearly renewal of storage space does not support the 5-year data retention requirement from the "Tri-Agency Open Access Policy on Publications" (34).
- Develop an automated dashboard to audit open scholarship practices, including open access publishing and data sharing, to promote trust in repository infrastructure and researcher practices (41).
- Develop policies and messaging that promote RDM and data sharing, and incentivizes institutions to develop local RDM policies (27, 62).
- The Alliance and other collaborators in the scholarly ecosystem should consider the engagement and influence of commercial publishers in further adopting open scholarship principles (47).
- Consider working with organizations to make their data discoverable and/or accessible in a secure environment, which would allow their data to be used for research (37, 77, 96).

4 Operations

4.1 Defining scope

In its development, the Alliance should consider the full spectrum of DRI funding and ensure that important funding streams that currently exist are not lost in the creation of the Alliance.

Community recommendations:

- The Alliance should clearly define the role it will play in supporting DRI, and who and how it will collaborate with other players (13).
- The Alliance should also consider the existing organizations that are merging and identify any funding gaps that exist so that the creation of the Alliance as an organization can fill (e.g., 07, 13, 23, 81, 83 and others).

4.2 Equity, diversity and inclusion

DRI continues to have significant challenges with EDI–in terms of users, disciplines, regions, institutions, and even Professional Support Staff. The Alliance can help materialize change through hiring, training, leadership, and opportunities (22, 46, 53, 54, 55, 33). EDI can also serve as a lens for assessment. If EDI is not improving, it points to missed opportunities and gaps that should be addressed (41).

4.3 Organizational structure

Authors noted the benefits of centralized infrastructure, including reduced duplication of effort, economies of scale, collaboration, and inclusion (08, 98). Examples of successes of a centralized infrastructure that can be expanded upon included the common scheduler, software stack, authentication, documentation, and support systems, among others (80). On the other hand, the downsides a centralized model were identified as poor support for users with unique needs, and reduction in personal contact and engagement with HQP and Professional Support Staff (80). Some authors described that the current model of a federated partnership (with regional partners) adopted by CCF is not very responsive to the researchers' needs, had difficulties in speaking with one voice to make adequate representation to government, and was not directly accountable and ultimately fomented divisions (14).

Other authors alluded that because decision making is currently a local function, different host sites develop different services. Moreover, when coordination does happen and the same service is deployed on multiple sites, it is usually deployed independently by different people, which results in different implementation decisions (93). One additional challenge of such system is that the site personnel are often employees of their local institution or university, and not Compute Canada, which leads to confusion about responsibilities (93).

In addition to funding shared infrastructure, a submission from a structural dynamics lab also noted that labs need funding to develop and deploy their own DRI, which is essential for innovation (10).

Community recommendations:

- Authors from the Global Water Futures research group recommended a tiered support model as one mechanism to meet basic DRI needs experienced by all researchers (15). Effectively they recommended that the Alliance provide a basic collaborative research platform to all researchers with low barriers to use, in terms of usability and cost. The example given was the Purdue University Research Repository (PURR) system at Purdue University, based on Hub Zero, which offers access through a tiered resourcing model. All researchers receive the base level of resources at no cost, researchers with grant-funded projects or other supports receive additional resources, and storage can expand at a reasonable cost.

- Authors from the Institute of Particle Physics and the Canadian Institute of Nuclear Physics highlighted that hardware systems and personnel could be structured by computing requirements, and allow researchers to engage with teams with specialized expertise (92). They proposed subgroups, such as data curation, archive and preservation, high-performance (Message Passing Interface, MPI-type) computing, high-throughput and high input/output data computing systems, and general-purpose systems.

4.3.1 Harmonizing policies among Host Sites

Multiple papers have noted a need for more consistent polices in Compute Canada Clusters, and that work to streamline and standardize operations should continue (67). All local policy decisions affect the ability of researchers to port work to different clusters. CFI treats each site as a separate project; each site's host institution owns the hardware and run the procurement process. Some authors advocated for a more transparent governance and incentive to cooperate and not compete, while accommodating the evident differences between sites (93, 51). Moving forwards, host sites require "stable and predictable funding for operating and a clear mandate to deliver services that are consistent across the national platform" (67).

Community recommendations:

- Need for integration of policy development at sites and with national teams. Consultation is often not done which leads to frustrating and lengthy development cycles integrating the necessary distributed computing platform infrastructures (21).

4.4 Community engagement and collaboration with regional stakeholders

The need for close collaboration with and support for regional stakeholders, particularly regional computing service providers and library associations was highlighted by research institutions, and in many cases the stakeholders in question (40, 46, 58, 36, 97). For instance, Atlantic Computational Excellence Network (ACENET) emphasized the need for sustained and bidirectional communication, as well as close integration of employees at different levels within the Alliance working groups to leverage their expertise (46). In their submission, the Université du Québec à Chicoutimi (UQAC) posed the question of whether Calcul Québec could act as the Alliance's representative in Québec, given their excellent history of service. At the very least, mandates need to be clarified between the Alliance and regional entities (56).

Community recommendations:

- Collaborative development of infrastructure with regions was also raised as a mechanism to leverage community investments and deliver scalable infrastructure, while also allowing for local differences. The collaborative model of Dataverse Canada is one successful example (94, 89, 08, 40).

- User groups for specific components of the Alliance's infrastructure could include members of the Alliance's team managing that component. In addition to building peer support and potential collaborations, it would also provide the Alliance direct input into services (54, 67).

4.5 International collaborations

International collaborations are essential to many fields of research, and the Alliance should enable connections with international counterparts to advance collaborative enterprises and shared standards (17, 21, 12, 92). The Alliance is also regarded to be in a good position to fund research groups for their participation in efforts to develop interoperability internationally, in research disciplines, and between research institutions (67).

The international nature of many research areas requires that resources integrate with the global infrastructure (03, 12, 17, 19, 30, 47, 71, 99), such as parallel computing that must be conducted through international collaborations. An internationally interconnected Canadian DRI would avoid single-point failures and would facilitate distributing types of computation among different DRI (e.g., map making for Atacama Cosmology Telescope (ACT) done on Niagara, KwaZulu-Natal, and NERSC clusters; 71).

Community recommendations:

- Collaborations with international partners often rely on reciprocal benefits. For instance, offering infrastructure to international communities and in return Canadian researchers receiving priority access to data or scientific instruments (e.g., the astronomical and particle physics communities). Mechanisms must be in place to support these connections, and the Alliance should continue to provide infrastructure access to international collaborators.
- The Alliance, at both management and operational levels, should consult with international funders (e.g., National Science Foundation; NSF) and organizations that run large computer systems (e.g., National Energy Research Scientific Computing Center (NERSC) and The European Organization for Nuclear Research; CERN) (14, 92).
- National infrastructure needs to enable international collaboration through adoption of shared standards or policy frameworks for interoperability.
- Coordination mechanisms like working groups and federations ensure that skilled personnel and technical infrastructure are aligned.

4.6 Identity management

The importance of identity management, two-factor authentication, and the importance of multi-discipline collaboration tools were often addressed in submissions, given that existing tools like Slack, GitLab, staff wikis and others have strict identity boundaries between those within the organization and those outside. This has been shown to prevent external collaborators from participating (21).

Community recommendations

- The Alliance should work with its member institutions to implement federated identity systems, likely under the CANARIE Canadian Access Federation (21).
- The Alliance should implement multi-factor authentication and other controls and tools to help safeguard research data and prevent security breaches (32).

4.7 Supporting big infrastructure projects

Several submissions emphasized the importance of supporting large-scale projects through Professional Staff capacity, and funding.

Community recommendations:

- The Alliance should act as a potential partner in the Canadian Square Kilometer Array (SKA) data centre among other Canadian stakeholders (e.g., CANARIE, the Canadian Astronomy Data Centre (CADC), the Canadian Advanced Network for Astronomical Research (CANFAR) and Canadian universities) (03).
- Collaborate with the CIHR-led Canadian human genome library project that will integrate human genome sequencing efforts across Canada and establish a "single point of entry" for access and analysis (61).
- Collaborate with the FNIGC on its vision for First Nations-led network of regional information governance centres will serve the information needs of First Nations communities across the country (98).

4.8 Professional development and employment security for Professional Support Staff

Many authors emphasized the importance of national Professional Support Staff, especially given their role in cross-institutional research environments (28, 46, 67, 101). Many papers credit the current support provided by CCF Professional Support Staff, but note the precariousness of their employment is a risk. Currently, many of the CCF staff are supported by short-term contracts, when long-term stability is needed to help recruit and retain HQP (17). This is especially the case in the digital humanities where salaries for system administrators and developers are not eligible expenses in regular grants (17, 54).

It was also highlighted that given the speed at which technologies and practices are changing, Professional Support Staff require constant training and professional development to remain up to date in skillset and expertise to support researchers in specific areas (27; see Education and Training).

Community recommendations:

- Professional Support Staff within the Alliance should be cross-trained in their skillsets to ensure avoid single points of failure (92).
- Professional Support Staff should maintain awareness of developments in DRI to anticipate needs and suggest platform enhancements or replacements as technology becomes availably (65).
- The Alliance should support Professional Support Staff for their continued development and provision of training opportunities, learning resources, and guidance materials in both official languages (27).
- Develop specific funding streams and/or agreements to cover the salaries of Professional Support Staff at all levels (i.e., national, regional, local) and ensure long-term employment security and sustainability (15, 54).

4.9 Resource Allocation Competition (RAC) process

Many authors expressed challenges and frustrations regarding the RAC process (2, 15, 17, 21, 71, 92, 105). There is currently no formal connection between the yearly RAC application process and allocation of resources, with the actual purchase of resources at the sites. Some identified that the RAC applications themselves are onerous and need to be frequently renewed. Others highlighted that the time scales of the RAC process do not match research timescales and there is currently no mechanism to apply for compute time outside of a normal RAC cycle. For example, many HPC initiatives require concentrated resources for short time periods, rather than a uniform share of resources distributed over the entire year. On the other end of the spectrum, some authors noted that many researchers operate multi-year projects, which require access to DRI over multiple years; even with the “fast track” option for two more years after initial application.

In addition to the actual allocation, some authors emphasized that the communication between the operations teams and the RAC committee ought to be improved to ensure that allocations match research requirements, as described above. Some authors also shared frustration with the RAC process, as some research projects have had their allocation suddenly reduced at mid-project due to uncertainties in the annual RAC competition. This is particularly frustrating when the project has already passed peer review and been awarded Tri-Council funding.

Finally, some authors shared their concerns about the RAC process and the challenges it poses to research programs that require allocations to meet the needs of a group-projects, or a large project with multiple distinct computational components. Allocations being tied to the Principal Investigator (PI) result in challenges for large group projects and projects spanning many years with turnover.

Community recommendations:

- Align allocation of resources with Tri-Council (see Alignment with Funder and Institutional Policies).
- Provide team allocations with a multi-year duration.
- Create a fast-track RAC stream for researchers needing short periods of computing (e.g., larger, massively parallel, and/or burst type) (02, 29, 105).

5 Support

5.1 Professional Support Staff

Many authors described support received from Professional Support Staff as equally important component of DRI as physical infrastructure, if not more so; with key roles played by Professional Support Staff within research institutions and nationally. "*Researchers look to IT services, libraries, research services offices, digital humanities/digital scholarship centres and similar, and other on-campus units to find technology and technological supports where possible; as a second recourse, they turn to national societies or federally-funded supports; and, as a last resort, to commercial products; or, perhaps worse, they never realize their project's potential.*" (22)

5.1.1 National

Several authors also identified the need for more subject-specific DRI expertise (18, 25, 67, 83). Human resources are needed to work with research groups to develop and operate their discipline-specific code instruments, data pipeline, etc. The Canadian ATLAS Tier-2 facilities, for example, are one successful model of embedding personnel with subject-matter expertise within CCF and host sites (99, 92). The Bioinformatics National Team also described resourcing challenges limiting their impact (25).

Researchers from the Humanities and Social Sciences (HSS) expressed need for additional support using software and computing (55, 18). Within HSS disciplines, where use of technology is less frequent, researchers have to relearn tools in every project (one suggestion given was Linux system support for setting up and running web servers), whereas access to dedicated Professional Support Staff would be more efficient and maintain "muscle memory" of researchers (29).

Community recommendations:

- One role suggested for the Alliance was to focus on services that maximize use of the entire DRI ecosystem, not only the Alliance -specific, by helping the researcher locate “pieces of the puzzle” and assemble them into a solution (13).
- Support for Research Platforms as a Service (PaaS) is proposed, whereby a central team of Professional Support Staff be created to support all the Research Platforms relying on the DRI (75).
- Cloud support via a national team to build national capacity is needed to address lack of IT and system administrators, privacy and security, and specialized software expertise for coding, and reduce duplication of technology efforts between projects and risk of inequity (104, 38).

5.1.2 Institutional

Many submissions referenced the important role played by Professional Support Staff distributed within institutions that provide in-person support. The roles played by libraries, in particular, was highlighted as critical in providing researchers with RDM support, training, and guidance–especially with the RDM policy from the Tri-Agency. Library-based Professional Support Staff have direct ties to their researcher communities and knowledge of institutional solutions. However, libraries often face staffing challenges related to capacity and professional development. Finally, advanced technical research data management support for researchers can often fall between the mandates of institutional libraries and IT–a gap that researchers often struggle to fill (55).

Community recommendations:

- Institutional Professional Support Staff expertise capacity development through national initiatives like Portage should continue (08, 20, 27, 28, 89, 101).
- The need to fund institutions to build human infrastructure, and support their professional development is important to their success and retention (8, 46, 67, 27).

5.2 Education and Training

Different approaches to training were highlighted as imperative to serve the needs of diverse communities, and their equally diverse forms of data. Further, this training should support a range of experience levels (53).

5.2.1 Inequities between regional and institutional training provision

Some submissions alluded to the significant differences in the provision of training and training resources. This was exemplified by differences in ARC training across universities and, even departments (46). As well there are significant differences in training opportunities provided by regional consortia (105).

Community recommendations:

- Support staff within institutions were also discussed as effective ways of providing training and nationwide coordination of user support and training (67).

5.2.2 Graduate training

Some authors identified the limited time graduate students have to ramp up and start research as an important limiting factor in the development of DRI-related skills (15).

Community recommendations:

- More agile training to facilitate students in understanding the characteristics of research platforms, and in mapping their research-specific infrastructure requirements (40).

5.2.3 Computing

Some authors mentioned that Compute Canada already provides the basic building blocks upon which researchers could build their own system (e.g., virtual machines, IP addresses, software etc.). Yet there still exists an important knowledge gap and a high barrier on how to use these tools (15; see Accessibility and usability of infrastructure). More advanced training on the use of ARC or "how to wisely use ARC" is also needed (105).

Community recommendations:

- Provide intermediate to advanced training on schedulers, parallel programming with MPI, openACC, version control, and remote parallel visualization of large datasets. Training on diagnostic tools to evaluate performance, communication, and read/write latency of massively parallel simulations (105).
- More discipline-specific training should be provided. For instance, offer thematic summer schools based on disciplines rather than tools (105).
- Calcul Québec (49) describes the tools and education needed and proposes this work could be led by the formation of a specialized team who could work with researchers to guide them in implementing practices.

5.2.4 Research Software (RS)

RS is an important part of modern research, and while large research groups have capacity to recruit developers, small groups with less capacity rely on postdocs and graduate students who are primarily self-taught.

Community recommendations:

- The Alliance's RS strategy should incorporate a focus on best practices in software development (67).

5.2.5 Research Data Management

Many researchers require training in RDM, such as data curation and storage, but also in understanding and navigating the range of data repositories available to them (02).

Community recommendations:

- Some authors suggested that further development of Data Management Plan (DMP) - related training resources is needed to ensure Canadian researchers are equipped to write effective DMPs (82).
- Researchers from the Université Laval promoted champions program (e.g., https://www.data.cam.ac.uk/intro-data-champions) to advance RDM engagement within institutions (14). These Data Champions could engage with research communities and play a mentor and trainer role.

5.2.6 Data Literacy

To meet Canada's research objectives increased attention to building data literacy skills was identified as paramount (45).

Community recommendations:

- Submissions identified a role for the Alliance to play in convening communities together to engage them for the advancement of DRI (e.g., Portage Network of Experts), including establishment of standards and best practices in various communities and fields of research, encouragement in adoption of data sharing, reuse, and other RDM practices, hosting discussions between researchers, practitioners, policy makers, among others (e.g., 46, 35, 54, 62, 71, 72, 27, 96, 40 and others).
- Engaging with the granting agencies and the Alliance Researcher Council to determine which disciplines require enhancements to their data literacy competencies.
- Assessing the need for a series of boot camps/summer schools/grants for graduate students, researchers, data stewards and research assistants who will be working with researchers.
- Lobbying universities to promote teaching data literacy competencies in all Faculties.

5.2.7 Indigenous Data Sovereignty

First Nations will need opportunities and capacity support, including funding for training to support the development of their technical capacity, as needed (36). On the other hand, Canadian data experts and leaders would benefit from Ownership, Control, Access, and Possession (OCAP®) training and cultural awareness training (102, 36).

Community recommendations:

- Promote the broader training on and adoption of OCAP® and CARE principles for the research community.

5.2.8 Training for Community Building

Forums for knowledge sharing were recommended to take advantage of the vast knowledge resources that are available for researchers (02, 24, 57, 83). Models such as FemTechNet were identified as excellent resources, even more so than mainstream tech sites such as StackOverflow, which disadvantages women (54).

Community recommendations:

- Links between universities, disciplinary associations and CCF support services should be enhanced through specialized training on subject-specific topics (105).
- Training opportunities and national infrastructure that supports a curation community of practice should be supported and will allow curators to network, share knowledge, expertise, tools and best practices (27).

5.3 Partnerships

Collaborative and coordinated support models were identified as one mechanism to defragment the current landscape (43, 27, 16, 15, 7, 102). The collaborative support model that Dataverse Canada is founded on is one example of a successful partnership between research institutions and national support, which could serve as a model for other shared services and infrastructure (e.g., curation, and preservation). Another example is the Ontario Primary Care Learning Network: Managing, cleaning, and analyzing primary care electronic medical records (EMRs) by Practice-Based Learning Networks that develop partnerships between researchers, clinicians, and data scientists (7).

Community recommendations:

- The Alliance should foster connections and support between academic institutions and other stakeholders of the wider research ecosystem (including Indigenous Peoples, disciplinary repositories, non-governmental organizations, and research groups).

5.4 Sustainable Funding

Several authors shared their concern for the lack of funding opportunities for existing projects that have ongoing operational needs like hiring or retaining HQP and Professional Support Staff, or undertaking project updates (20, 22, 102, 106). Existing opportunities favour employing students, which provides a significant benefit for their training, but does not provide the opportunity to capitalize on developments and their experience acquired over time (10, 16).

6 Research Data Management

6.1 Data Management and Open Science

6.1.1 Interoperability

6.1.1.1 Data/Metadata

The need for interoperable workflows involving data and associated platforms used in the research process was frequently discussed in submissions related to Open Science/ RDM.

Interoperability at the data-level is a challenge for many research communities, especially the Humanities, Biological Sciences and for communities working with qualitative data, resulting in many digital projects and repositories in these disciplines using their own unique ways of organizing and describing data (01, 04, 52, 57, 83, 94). Metadata production is often left out of project budgets. However, this way of conducting research increases costs to researchers and decreases accessibility and interoperability of data (01). Cost benefit analyses are needed to demonstrate how an investment upfront in metadata has downstream benefits to encourage development and adoption of standards in these research communities (24). Even where metadata standards and ontologies exist, specialized tools for capturing and producing metadata, and managing workflows are needed (04, 61, 52, 24, 83).

Automated creation of metadata was also discussed as an area for further support; so was automated indexing and data linkage (74, 83). Processes to set up data linkage with other datasets can be time consuming and resource intensive. Nimbler processes and procedures are needed (7). Privacy-protected linkage is needed to connect provincial data (e.g., in health and education) with federal data (e.g., Census, personal income, immigration; 96). More linkable administrative records should be made accessible to researchers (96; see Sharing PII and health data).

Greater support is needed to enable interoperability of qualitative research data. Currently metadata standards are designed primarily for quantitative data, and they can be potentially extended to accommodate qualitative data, leading to a need for more specific standards and repositories (94).

6.1.2 Platforms

Research platforms play an important role in DRI by enabling interoperability of data, and the Alliance needs to support funding calls for new platforms (*researcher-centric innovations*) and improvements to existing platforms (*researcher-led feature development*) (18). To connect platforms, a set of secure and reliable mid-level access tools (e.g., APIs, Single Sign-on (SSO)) that will enable easy interconnections between the platforms and the underlying ARC infrastructures are needed (18).

Coordinating technology developments between research platforms, data repositories, preservation processing systems, and long-term storage are needed to ensure interoperability (28). Digital preservation systems do not work well with content management systems, requiring parallel systems duplicates the effort of ingesting and curating digital assets (18).

Connections between platforms to link research data and other types of data are also needed, particularly for health and medicine. HDRN is leading this work in Canada to bring research data together with clinical records, provincial billings, diagnostic imaging, referral reports and laboratory information (5).

The need for interoperability between data repositories, institutional repositories, open educational resources, and other platforms supporting open scholarship was also raised by Canadian Association of Research Libraries (CARL-ABRC) (08).

A related need is interoperable Research Information Management (RIM) systems for improved data collection and comparisons across institutions, with benefits to research graph development (08, 09, 32).

6.2 Standards

The need to support the development of standards for both data and tools was often discussed. For instance, existing repositories in biosciences use different methods for submitting, describing, and retrieving data objects (04, 52). Some disciplines have their own standards that national DRI needs to be able to support to enable interoperability. (e.g., GA4GH, International

Image Interoperability Framework (IIIF), and other standards mentioned) (1, 12). For health data, provincial organizations have their own procedures for data quality and standardization (e.g., The Institute for Clinical Evaluative Sciences (ICES) in Ontario and PopData BC), which need to be harmonized (e.g., work of the HDRN) (5). For further information see Sharing PII and health data.

Community recommendations:

- Standards-based certification efforts for data repositories and preservation infrastructure are needed to support and connect trusted infrastructure (28).
- Development/improvement of standardized ontologies that define vocabularies and harmonize data from different systems are needed to provide a common way to describe and classify datasets (04, 12, 83).
- Standards are also needed for governance, policies, and architecture of systems and institutions (12, 32, 52, 83).
- Industry-standard security frameworks are also needed (e.g., authentication, authorization, intrusion prevention, digitally signed applications) (12, 83).

6.3 Persistent identifiers

The submission from DataCite Canada and the Canadian Open Researcher and Contributor ID (ORCID) Consortium described the need for critical mass uptake of persistent identifiers (PIDs) to fulfill their intended purpose (89, 90). DataCite Canada and ORCID-CA currently provide many Canadian research institutions access to two key PIDs, ORCID and DataCite Digital Object Identifiers (DOIs) (6).

Other authors suggested that the current identifier systems all rely on central registries, which pose issues of interoperability, fragmentation and trust, as well as costs to maintain. They suggest that rather than maintaining registries or support registries, the Alliance could support infrastructure (e.g., tools, policies, software, knowledge, training) that empowers organizations and communities to run and maintain their own PID systems. They further suggest starting with a proof of concept and lists existing technologies that could be leveraged (26).

Community recommendations:

- The Alliance should provide sustainable for the development of persistent identifiers and promote their adoption through their integration in all managed/funded Alliance infrastructure.
- A national PID strategy is needed to investigate adoption of emerging PIDs (e.g., Research Activity Identifier (RAiD), GrantID) (6, 90).
- The Alliance should incorporate all digital resource identifiers into their workflow and products (26).
- Create "*an ecosystem of Decentralized Resource Identifiers for research that are easier to create and maintain, lower cost, more flexible, and eliminate the danger of single points of failure*" (26).

6.4 Data Management Plans

The DMP Assistant in concert with the many resources provided by the Portage Network, gives researchers the tools to develop a solid DMP. While these resources offer guidance on best practices, it is necessary for the digital infrastructure to be in place to enact them and to facilitate researchers' compliance with requirements forthcoming requirements from the Tri-Agency (82).

Researchers also receive directive on how to manage their data from the Research Ethics Board process and need infrastructure and support for those requirements.

The Portage DMP Assistant supports improved data management practices and responds to the requirements of the new Tri-Council RDM policy. However, it is not enough on its own and must be supported by expert guidance and resources (65, 103).

The paper submitted on DMPs identified an increasing demand for planning support, the need for diverse templates and exemplars for subject areas and communities, and multilingual support (82).

Community recommendations:

- Guidance on evaluating DMPs was identified as a significant gap (40) with the development of national standards recommended as one option (32).
- There is a need for the Alliance to build on support for DMPs and provide a national standard to reference (32).
- There is an opportunity for greater alignment between REB applications and DMPs (15).
- The Alliance also has an opportunity to incorporate DMPs into the existing CCF resource allocation processes (82; see Resource Allocation Competition).

6.5 Research Data Management Support During Active Research

A greater connection between RDM and ARC was identified as lacking. Currently, RDM is left to the researchers to manage on Compute Canada's infrastructure, when they often have limited knowledge in this area. The paper from Calcul Québec (49) describes the tools and education that are needed and proposes this work could lead to the formation of a specialized team who could work with researchers to guide them in implementing RDM practices in ARC. The authors shared the following examples:

- *"Help researchers structure their data, improving their efficiency and their ability to work in a team;*
- *Anticipate the creation of metadata, thus facilitating their preservation and reuse in the future;*
- *Anticipate the transfer of intellectual property between students and principal investigator at the end of the study cycle;*

- *Avoid cluttering file systems with unpurged temporary files;*
- *Avoid duplicating master datasets downloaded multiple times by different people to a local directory rather than a shared central repository."*

Community recommendations:

- More support for RDM is needed during the active research process, and the Alliance could provide tools and support that bridge the existing gap between the DMP and the long-term repository (15). Areas mentioned for further support across papers include structuring data, file management and version control, metadata creation/entry, connecting storage types.
- The report submitted on national curation (27) also expressed that the potential for data reusability and study reproducibility would be maximized when curation support is integrated into all research support processes, including the active research phase.

6.5.1 Data Curation

Research libraries were often cited as key actors supporting data curation in higher education institutions, who require coordinated, national support in leading this work (20, 27, 36, 40, 67, 89).

Over 40 unique data repositories are referenced across papers. The development of standards, ontologies, and policies for interoperable data and platforms underpin these infrastructures. However, their success relies on skilled Professional Support Staff (e.g., curators) to develop and apply such standards, ontologies, and policies to datasets, in order to harmonize and integrate infrastructures.

Community recommendations:

- Specific funding streams are needed to support Professional Support Staff to develop and maintain these repositories and underlying infrastructures, and curate the data they contain (67).
- The work associated with curation is time consuming but there is limited support (e.g., time, training, compensation) (57, 94). The Alliance is uniquely positioned to enhance the support of national curation (54).
- Recommendations from the report submitted on national curation emphasize partnership and collaboration to support curation. For instance, that the Alliance should support mechanisms for inter-institution collaboration for curators and other Professional Support Staff to share resources, best practices, etc. (27).

6.6 Management of sensitive data

Several authors confirmed that the lack of secure and standardized solutions for managing sensitive data is a major challenge and requires a federated solution to address differences across provinces, territories, and Indigenous Peoples in privacy regulations and health data standards (see Sharing PII and health data). Such effort must be a collaboration across "pillars" and efforts under way should be supported (e.g., CanDIG, HDRN, CRDCN) (67, 96).

A coordinated approach to collection of data related to race, ethnicity, and other social determinants of health was identified as lacking, and one must be developed with deep involvement of members of affected communities (e.g., policy frameworks, standards) (35, 79).

Several authors reiterated that the ethical issues surrounding sensitive data–and some data sets–should not be made available for reuse (94). Yet, the infrastructures where protocols for storage and access can be co-created and negotiated with data owners are needed (e.g., data from Indigenous Peoples, personal health data) (94, 102).

Some data suffer from limited discovery and access due to size, resourcing, and technological conditions. Repositories require features to support this data (e.g., access control, secure transfer, licensing; see Identity management), and communities need to develop related metadata, reporting standards, and policies to support discovery.

Research institutions are lacking access to resources to manage confidential information within DRI, and that comply with relevant legislative requirements. For example, tools for securing data transfer, and support with key management practices (e.g., anonymization and de-identification) (32, 40).

Community recommendations:

- There are few resources to guide researchers in planning for the secondary use of data beyond the TCPS2. While individual REBs may have the resources to support researchers, practices are inconsistent. Better tools to support both secondary use of sensitive data are needed, including tools for encryption, and secure environments for regulating data access (39).

6.7 Data discovery

Progress made by Portage *via* FRDR to support data discovery through a national catalogue and discovery layer was acknowledged, but there was also recognition that more attention is needed and there is a role for greater software integration to support data discovery.

There is a need for improved registries of specific types of data that are currently challenging to locate: access-limited data (37) (including health data (04, 05, 35)), geospatial and cartographic data (60), historical data collections (43), and web archives (89).

Many of these resources are governed by government agencies such as Library and Archives Canada, the Treasury Board of Canada Secretariat, Statistics Canada, Natural Resources Canada, and similar organizations at the provincial level, with whom the Alliance should form partnerships.

An inventory that is interoperable with other (national and international) registries and catalogues is needed to facilitate integration of Canadian research data into the Global Open Research Commons (44).

6.8 Digitization

There is a pressing need to document historical information, cultural heritage and knowledge, especially present within Indigenous Peoples, requiring national support to enable the collection and preservation of languages and oral histories in accordance with their Nation's governance (29, 102). Library and Archives Canada's (LAC) Indigenous Documentary Heritage plan and B.C.'s Indigenization initiative are examples (102).

Community recommendations:

- Infrastructure should allow for collaboration with cultural institutions and Indigenous Peoples to support the digitization of cultural heritage (16, 102).

6.9 Long-term RDM support

The submission from the Portage Digital Preservation Expert Group and the Canadian Association of Research Libraries Digital Preservation Working Group explains the mandate of Compute Canada has no allocation for providing supporting services for research data beyond the life of a specific, funded research project. This approach is short-sighted and incompatible with needs of data-intensive research (28).

7 Research Software

7.1 Sustainable RS strategy

The importance of RS as a scientific output was often cited. Several authors, however, described that local resources for its development are needed, especially in the form of training and support. Further, they highlighted that finding collaborators or consultants can be challenging (54), and that relying on short term developers and grad students poses issues for the maintenance and sustainability of RS (63, 67, 38).

Community recommendations:

- Many submissions suggest funding developers to work within research groups (17, 38, 92, others). They recommend that the Alliance should continue CANARIE's Local Research Software Development Fund. This is time-sensitive for existing teams, which will run out of funding in 2021.
- Authors from the sub-atomic physics community note that CANARIE's focus for funding software has shifted towards from impact to value-driven measures and should be revisited. Authors recommend that funds in these areas should be focused on the highest priority projects that have significant international impact based on peer review.
- The development of new software should be encouraged through pilot projects rather than trying to predict success of individual tools (92).
- The development of open-source tools should be supported to alleviate roadblocks related to proprietary data formats and specific license restrictions (10, 84).

7.2 Software underlying ARC

The report from University of Victoria Computing advocates for funding for research software to support the software components that form the underlying shared infrastructure and platforms of ARC systems, on which all ARC user applications run. For example, host sites use a range of open-source software like Simple Linux Utility for Resource Management (SLURM) for batch job scheduling, Lustre as a shared cluster filesystem, and CernVM-File System (CVMFS) for software distribution across sites.

Community recommendations:

- The Alliance should fund one or more software development positions dedicated to the open-source software projects and components used in CCF's infrastructure.

7.3 Research Platforms

All research communities from the sciences to the humanities have called for the Alliance to extend support to research platforms, supporting the concept of cloud-based Platform-as-a-Service (PaaS) (54, 75), which is now a part of the "*democratisation (sic) of HPC*" as more researchers are relying on platforms to conduct their research (75). For further information see Computing needs by research discipline.

Some authors described the need for cloud-based services that provide sandboxes for experimentation (16) and *dedicated Test and Development* platforms to test upgrades or on-board new experiments without disrupting the production systems (21). Support for Kubernetes (55, 54), Git Lab, web-based Customer Relationship Management (CRM) (55), HubZero (15) are some of the examples listed. Exploratory pilots and/or engagement with existing NSF pilots is one suggestion for the Alliance to support innovation in this area (14).

Community recommendations:

- A national repository was often recommended to make publicly funded software and source code more accessible (56, 100; see Awareness of existing infrastructures). An example given is l'INRIA in France, or a collaboration with Software Heritage (56). The CANARIE portal is a good step in this direction but authors from the Faculty of Science at UQAM note that it doesn't support important aspects like hosting or documentation, like GitHub (100).
- Given the expressed science need for data collections to cut across sub-disciplines (such as X-ray, optical, optical, infrared and radio astronomy) a single astronomy discipline aware science portal that enables use of the full spectrum of this data is needed (30).

7.4 Software supporting RDM

The burden of repetitive tasks, including curation and data discovery, ought to be transferred to information systems through use of APIs (01, 04, 24, 74, 84, others). Accompanying open-source software for curating, submitting, storing and discovering research data from different disciplines must be developed to achieve this, based on standards formalized by international organizations (e.g., GA4GH) (04, others). While many standards have and are emerging, the lack of tools is a significant barrier (24).

8 Advanced Research Computing

8.1 Renewing Investments in ARC

The importance of computing infrastructure was a consistent topic across submissions from academic researchers. Many authors described that for Canadian researchers to remain competitive, sustained investment in computational capability is needed. Moreover, this investment in ARC infrastructure spending ought to accommodate the expected increase in workloads from Canadian researchers (see Computing needs by research discipline). For example, the submission from the Simons Observatory confirmed that their annual requirement is expected to increase to roughly 30,000 CPU core-years by 2029, or 18 times higher than the anticipated cycles needed for ACT processing in 2021–22 (71). Similarly, the survey conducted by Genome Canada identified that the existing infrastructure, although incredibly valuable, does not offer sufficient computing power for the processing and analysis of genomics data (61).

A paper from authors across disciplines relying on parallel super computers presents a case for an ambitious renewal and expansion path for the large-parallel simulation capability in the Canadian DRI ecosystem, to build on existing progress, and to expand the associated specialised human resources required for scientific discovery (19).

Community recommendations:

- The Alliance should begin a program with sustained investment in computational capability (17).
- Invest in real time data processing on computing and Graphical Processing Unit (GPU) resources with timely feedback during experimental procedures, connected to stable data streaming and sensor networks (14).

8.2 24/7 services

Some authors emphasized the need for continuous support and services for national platforms. Specifically, they describe that certain projects like those that engage with international research teams, require jobs running 24/7 and therefore require 24/7 support (21, 63, 92).

Community recommendation:

- The Alliance needs to ensure that the facilities are designed and managed to be fully operational around the clock, and to ensure that scheduled downtimes do not result in loss of access to resources for any user (92).

8.3 ARC-related Communications

Some generic email-based ticketing systems are inadequate to provide the necessary information to the right person at the right time, and thus causes delays that compound confusion about who is the responsible contact.

Community recommendation:

- Adopt communications tools such as a transparent issue-tracking system (54).

8.4 Commercial Cloud

Several authors emphasized the importance of Commercial Cloud services for researchers, and that the Alliance would compete with them should it provide similar short-term services (56). Authors commented on researchers' need for flexible access without lengthy proposals, and that many researchers use commercial cloud for convenience (direct access, ease of use)–

especially when such services are not met by available national DRI (e.g., collaboration, allocations), or because they are not aware of local/regional/national options (20, 61).

Other authors discuss the barriers that exist in accessing cloud services for researchers, including privacy and security risks (see Risks below), issues with funding models not applying to operational expenditures, lack of dedicated resources to support the use of cloud, among others (107). Current grant regulations, specifically, don't allow researchers to pay for commercial cloud. "A hybrid approach would be the most effective mechanism for maximizing the benefit of the Alliance's investment in the Compute Canada Cloud (CCC) with the sophisticated Cloud needs of Canadian researchers" (50).

Finally, authors from the University of Victoria describe the risk for inequity in the cloud computing provision. Certain areas of research/smaller projects in emerging fields have not been provided with the same level of support that other areas have. Moreover, they allude that the demand for cloud resources currently exceeds existing national capacity, which leads researchers turning to commercial cloud solutions (104). The same authors also discuss the inaccessibility of Cloud Infrastructure-as-a-Service (IaaS) because it requires too much technical skills that small research groups cannot afford to develop (see Accessibility and usability of infrastructure). Moving to Platform-as-a-Service (PaaS) would lower the barrier to entry and let researchers focus on their research (104).

Community recommendations:

- A hybrid approach to supporting access to commercial cloud is suggested (50). For example, use for peak demands that exceed existing resources (92).
- Genome Canada recommends the creation of a Canadian genomics digital platform that blends private-sector flexibility with public sector subsidized resources and oversight (61).
- The Alliance could support research institutions by buying resources from commercial cloud and providing on-ramp to access services for Canadian institutions and researchers (02).
- Develop a short-term service that is competitive with the commercial cloud. The keys to this service are flexibility and accessibility. The Alliance could likely offer these services with more competitive pricing.

8.4.1 Risks related to commercial clouds

Supporting commercial cloud is challenging in the current funding model, which funds research infrastructure through capital investments, rather than operational expenditures (107).

There is also the risk that Commercial Cloud providers be motivated by high profits, which harms shared academic efforts (51). Therefore, most of the current barriers to using cloud resources are based on institutional concerns surrounding data privacy and unsustainable costs for large volumes of data (86, 61). For further information see Private sector influences.

Community recommendations:

- In engaging commercial cloud, the Alliance should address concerns around privacy and lack of support through developing a standard national approach to adoption through a Research Data Management (RDM) framework that considers the entire data lifecycle and expertise needed in each phase (107).
- Secure Canada's research data by harnessing shared cybersecurity services (88).

8.5 Computing needs by research discipline

8.5.1 Artificial Intelligence (AI) / Machine Learning (ML)

The 2017 Pan-Canadian AI Strategy has grown computational needs through recruiting more AI researchers to Canada. The submission from Vector Institute estimates that the AI demand for GPU computing is comparable to the entire GPU capability at Compute Canada federation facilities today (48). Meanwhile, some authors note that access to more GPUs through the national platform is needed to support researchers (not affiliated with one of the AI institutes) (103, 32, 61). In terms of operations, one challenge concerns the timescales with which data centres budget for upgrades, which is inconsistent with the speed of technological progress, especially in areas related to AI. Greater flexibility is needed to upgrade to the latest technology (48). Access to tensor processing units (TPUs) will be expected by the community to support AI growth (32).

Community recommendations:

- Infrastructure across provinces to support uptake of AI/ML approaches in clinical/health research was identified as a priority area for capacity development (12, 85).
- Vector Institute recommends that the Alliance increases ARC infrastructure spending to accommodate the expected increase in AI research/need; it is also suggested that that Alliance establish a joint governance model for operations and systems capabilities between regional computing centres and AI institutes (48).

8.5.2 Social Sciences and Humanities

The Centre de recherche interuniversitaire sur les humanités numériques notes that most often the research conducted in the Humanities involves needs in web hosting of specialized applications, software developments, and data preservation services, and less often access to GPUs/CPUs (16).

Community recommendations:

- Provide policy leadership and funding (e.g., through pilot programs) to catalyze more rapid innovation in the data-intensive research ecosystem, e.g., enabling cross-linking of (often sensitive) datasets that also require increased compute power (96).

- Provide an agile (cloud based) technological platform that responds to the variety workflows/pipelines used by humanists in their work coupled with on-the-ground human support (including code and web application development), with flexibility in stitching together services (20). This should include out of the box provision of DRI and related support for common development and scholarly communication tools, such as a Kubernetes cluster, GitLab instance, or CRM-backed website, rather than leaving individual research teams with the overhead and inefficiencies of installing and patching these systems themselves (55).
- Provide increased staff support dedicated to supporting advanced research computing and the development of new applications in the Humanities, including in the regional ARC facilities (18).

8.5.3 Health/Biomedical Data

8.5.3.1 Federated Approach

A survey of Genome Canada's users reveals a patchwork of non-interoperable computing infrastructures: *"Barriers to accessing a shared digital research infrastructure have resulted in the establishment of independent local computing infrastructures that are not interoperable. Our consultation revealed multiple examples of project data that span lab servers, institutional high-performance-computing clusters, regional academic clusters and multiple private clouds."*

Collaborative computing environments are needed to run complex operations on petabytes of sensitive genomic and clinical data (12). Given legislative and administrative needs of this type of data, increased capacity to perform distributed analyses on datasets that themselves do not move across jurisdictional boundaries is required (vs. requiring that datasets be brought together in a single location) (04, 35; see Sharing PII and health data).

Community recommendations:

- Develop a national-federated approach to unite expertise and data access across Canadian healthcare jurisdictions. Governance, architecture, ontologies, tools for interoperability are needed to tie all participating sites together into a single coherent network. For instance, distributed health and genomics data can be realized by bringing together separate initiatives (HPC4Health in Ontario, SD4Health, All For One, CanDIG, and Digital Health and Discovery Platform (DHDP)) with a single user access point and making them internationally operational by using standards set by the international discipline specific community of practice (GA4GH). Recommended models for federating data across repositories include the National Institutes of Health Data Commons (84).
- Develop a national strategy to address both present and future needs for PHI-compliant computing, storage, and sharing. To stay relevant and competitive internationally, this strategy will further need to fully embrace cloud computing and storage (86).

8.5.3.2 Data Trust Environments

Data trust environments for ARC were discussed by researchers in health fields, describing the need for infrastructure and personnel (data managers and system administrators) to support research on sensitive data through entitlements and encryption rather than restricted or isolated environments (85, 38).

Community recommendations:

- A secure platform for analyzing and storing sensitive data is needed- this platform must meet technical and legal requirements that require data to be stored on campus or within the provincial jurisdiction (14).
- The massive volume of sensitive health data generated today requires software and computer technology beyond what is available at individual centers across Canada (12).

8.6 Storage

Similarly to the request for an increase in investment in computing infrastructure, many researchers from across disciplines emphasized their need for more and faster storage solutions. For example, the submission from iReceptor identified the particular need for storage that can be easily accessed by compute nodes for processing large data sets (13). The submission from Genome Canada emphasized the high cost of long-term data storage. They describe that these costs pose a significant barrier to innovation by limiting the sustainability of the data assets and curtailing their useful lifetime (61).

Some authors addressed current issues with file management systems in national platforms, and their detrimental effects for researchers. For example, data redundancy across clusters, such as large reference datasets as the National Center for Biotechnology Information's (NCBI) Basic Local Alignment Search Tool (BLAST) non-redundant (NR) protein dataset cause overcrowding of the systems and affect user experience. The implications of this overcrowding of scratch systems is system volatility (25).

Community recommendations:

- Invest more resources for archival beyond the duration of the initial project to allow for secondary and tertiary mining of the (e.g., genomic) data such that initial investments continue to create returns in terms of new discoveries and solutions (61, 71).
- Invest in storage and compute capacity accessed via domain specific science platforms- the existing capacity within CC's centres will be exhausted on very short timescales (30).
- Establish protocols and policies to avoid cluttering the active temporary storage systems with data that should have been purged (49).

9 Position Papers

1. "Good Things Come in Small Packets": How (Inter)national Digital Research Infrastructure can support "Small Data" Humanities and Cultural Heritage research by Daniel Paul O'Donnel
2. NDRIO White Paper: Bridging the Accessibility Gap by the University of Waterloo
3. A CANADIAN SQUARE KILOMETRE ARRAY REGIONAL CENTRE by Kristine Spekkens, Erik Rosolowsky, Séverin Gaudet, Michael Rupen and the Association of Canadian Universities for Research in Astronomy
4. A Standards-Based Digital Infrastructure for Secure Sharing of Human Biomedical Research Data by Lincoln Stein on behalf of Ontario Institute for Cancer Research
5. Digital Health Research: Opportunities for Improving the Health of Canadians by Diabetes Action Canada
6. Persistent Identifiers in Canada: ORCID-CA and DataCite Canada by John Aspler on behalf of ORCID Canada Consortium, DataCite Canada Consortium, and the Canadian Persistent Identifier Committee
7. Ontario Primary Care Learning Network: ORACLE; EMR data to support research, quality improvement and innovation by the Ontario Primary Care Learning Network: ORACLE
8. Future-Proofing the Canadian DRI/RDM Ecosystem through Openness and Collaboration by Canadian Association of Research Libraries
9. Research Information Management Integration for Canada by Eugene Barsky
10. Digital research infrastructure of two research laboratories in the field of structural dynamics by the Laboratory for Acoustics and Vibration Analysis at Polytechnique Montréal and the Structural Dynamics and Vibration Laboratory at McGill University.
11. Canada's Health Workforce Digital Research Infrastructure Ecosystem: Building the Foundation for Canadian Health Workforce Science & Informed Decision Making by the Canadian Health Workforce Network
12. Digital research infrastructure to support federated computing on large scale biomedical datasets by Guillaume Bourque, Michael Brudno, Steven JM Jones
13. iReceptor – A case study in the challenges/opportunities in Canadian DRI by Felix Breden and Brian Corrie

14. Considerations from researchers from Université Laval by Jacques Corbeil, Francois Laviolette and Florent Parent
15. Gaps and Opportunities for NDRIO Support of Research Data Management: Insights from the Global Water Futures Data Managers Global Water Futures Data Managers by K. Dukacz and collaborators
16. Développer une infrastructure de services numériques pour les Humanités numériques canadiennes : livre blanc pour la NOIRN by the Centre de recherche interuniversitaire sur les humanités numériques
17. Digital Research Infrastructure for Canadian Astronomy by Catherine Lovekin, Pauline Barmby, J.J. Kavelaars, Adrian Liu, Erik Rosolowsky (U of Alberta), Kristine Spekkens, James Wadsley
18. Gaps in Digital Research Infrastructure for Canadian Digital Humanities Researchers by Lawrence Evalyn, Elizabeth Parke, Patrick Keilty, Elspeth Brown
19. Large-Parallel Supercomputer Simulations – Frontiers in Canadian Research by Rodrigo Fernández and 16 supercomputing researchers
20. Canada's Future DRI Ecosystem for Humanities & Social Sciences (HSS) Canadian Society of Digital Humanities by Geoffrey Rockwell
21. Submission to NDRIO's Call for Position Papers on Canada's Future DRI Ecosystem from the Compute Canada Federation Subatomic Physics National Team by Leslie Groer on behalf of the Compute Canada Federation Subatomic Physics National Team
22. All Researchers Use Digital Resources: On Campus Support, Grants, Labs, and Equity by Laura Estill on behalf of St Francis Xavier University Digital Humanities Centre
23. The need for expanding spatial data capabilities: Challenges and opportunities by Dr. Ines Hessler, ACENET
24. Empowering Information Systems and Fostering Metadata Driven Data Management by Pascal Heus
25. Canada's Future Digital Research Infrastructure (DRI) Ecosystem: A perspective from the Bioinformatics National Team (BNT) Bioinformatics National Team
26. The opportunities of Decentralized Resource Identifiers in the research landscape by Carly Huitema and collaborators
27. Closing Canada's "curation gap": A national approach by the Portage Curation Expert Group
28. Digital preservation and NDRIO: a white paper by the Portage–Preservation Expert Group and the Canadian Association of Research Libraries–Digital Preservation Working Group

29. Indigenous Language Technologies and Online Resources: Algonquian Dictionaries Project and Algonqian Linguistic Atlas by Marie-Odile Junker and Delasie Torkornoo
30. Digital Research Infrastructure in Astronomy Herzbeg Astronomy and Astrophysics Research Centre by Dr. JJ Kavelaars
31. The Needs of Canada's Future Digital Research Infrastructure Ecosystem by the University of Ottawa
32. Beyond Virtual or Physical Environments: Building a Research Metaverse A White Paper for NDRIO's Canadian Digital Research Needs Assessment by Paul De Decker and Stephany Peterson
33. Toward Petabyte Scale Open Neuroscience: UBC Dynamic Brain Circuits Research Excellence Cluster Experience by the UBC Dynamic Brain Circuits in Health and Disease Research Excellence Cluster
34. Population Health Digital Research Infrastructure: Building Up and Out from a Learning Collaborative Network by the Health Data Research Network Canada
35. Regional Academic Library Consortia as key partners in building sustainable, responsive Digital Research Infrastructure by the Regional Academic Library Consortia
36. Improving the Discovery of Access-Limited Data by Kevin Read, Amber Leahey, Sarah Rutley, Victoria Smith, Kelly Stathis
37. On the Need for Local Research Software Development Funding by Andrew Schoenrock
38. Strengthening Capacity to Manage, Secure and Protect Sensitive Research Data by the Portage Expert Group on Sensitive Data
39. Towards an Ideal Digital Research Infrastructure (DRI) by Robyn Stobbs and Vivekanandan Kumar
40. Creating an Audit and Feedback Loop to Monitor and Enhance Data Sharing in Canada by Felicity Tayler, David Moher, Kelly Cobey
41. Software support for ARC infrastructure by Ryan Taylor
42. Ensuring access to Canada's core historical data collections in the social sciences and humanities: possibilities emerging from the Canadian Census Discovery Partnership by Leanne Trimble and collaborators
43. Canadian Data Repository Inventory Peter Webster (St Mary U) and collaborators
44. A Proposal to Strengthen Data Literacy in the Canada Research Data Ecosystem by Ernie Boyko and Wendy Watkins
45. ACENET's Research Directorate White Paper for NDRIO by Andrew Rutenberg on behalf of the ACENET Research Directorate

46. Canada's Path to a Global Open Research Commons by Karen Payne, David Castle, and Mark Leggott
47. Canada's Future DRI Ecosystem: AI Research Needs by Ron Bodkin
48. Clés pour une Infrastructure de recherche numérique (IRN) innovante by Calcul Québec
50. The Future of Cloud as DRI by Deborah Stacey on behalf of the Linked Infrastructure for Networked Cultural Scholarship project
51. Comments for NDRIO by Mark Hahn
52. Individuals, communities and population health digital research infrastructure: incorporating primary care electronic medical record data as part of a primary health care information system by Sabrina T. Wong on behalf of the Canadian Primary Care Sentinel Surveillance Network
53. CSEE/CIEE Submission on the New Digital Research Infrastructure Organization by Stephen Heard and Diane Srivastava on behalf of the CSEE and CIEE
54. Interface Matters Canadian Writing Research Collaboratory by Jeffrey Antoniuk and Susan Brown
55. HQP Pathways: Engaging the Canada's Different Disciplinary Models for HQP Training and Funding to Facilitate DRI Uptake in Canada by Ray Siemens and Alyssa Arbuckle on behalf of the Digital Humanities Summer Institute Directorial Group
56. L'Infrastructure Numérique en Recherche Décanat de la Recherche et de la Création by the Université du Québec à Chicoutimi
57. Animal Behaviour in the 21st Century: A White Paper on DRI for Animal Behaviour Research in Canada Animal Behaviour Research in Canada – Dr. Laura Adams (Langara College) and collaborators
58. Écosystème D'infrastructure De Recherche Numérique Pour Les Données De La Recherche Vision Du Sous-Comité Des Bibliothèques Du BCI
59. Federating innovation in Digital Research Infrastructure accessibility Julie faure-Lacroix, Félix-Antoine Fortin, Florent Parent, Darren Boss, Drew Leske
60. Enabling broad reuse of Canada's geospatial data and digitized cartographic materials Association of the Canadian Map Libraries and Archives by Jay Brodeur and team
61. Genomics Digital Research Infrastructure to Support Canadian Competitiveness by Genome Canada
62. Role for NDRIO in advancing open data sharing and the use of domain-specific repositories by Carolyn DuBois and collaborators

63. Steps to Success in Ensuring DRI Engages and Mobilizes Humanities and Social Science by Research Ray Siemens and Alyssa Arbuckle on behalf of the Implementing New Knowledge Environments Partnership

64. Évaluation canadienne des besoins en matière de recherche numérique – Livre blanc de l'École nationale d'administration publique, l'Institut national de la recherche scientifique et l'Université TÉLUQ École nationale d'administration publique, Institut national de la recherche scientifique, Université TÉLUQ

65. University of Alberta response to NDRIO call for CDN DRI needs assessment white papers University of Alberta

66. BRISER LES SILOS DE DONNÉES ET DE SAVOIRS Un enjeu scientifique d'inclusion sociale et culturelle Université du Québec à Montréal avec la collaboration de Calcul Québec

67. Response to NDRIO Call for Position Papers on Canada's Future DRI Ecosystem McGill University

68. New Digital Research Infrastructure Organization White Paper Submission by Svetlana Komarova by the Faculty of Dentistry

69. White Paper: Open Database Support and Curation by Nicholas Probart, Gary Bader, Lincoln Stein and Michael Brudno

70. NDRIO Cartik Sharma Proposal by Cartik Sharma

71. NDRIO White Paper: Envisioning Digital Research Infrastructure for the Simons Observatory by Adam Hincks and collaborators on behalf of The Simons Observatory

72. NDRIO: Limit Corporate Influence, Maximize Public Involvement and Accountability by Catherine Stinson and collaborators

73. A stronger Canada through digital infrastructures that support primary care research and practice improvement by the College of Family Physicians of Canada / Le Collège des médecins de famile du Canada

74. Standardized, linkable, analysis-ready environmental data for understanding and preventing disease and building resilience to climate change by Jeffrey R. Broo, Dany Doiron and Eleanor Setton

75. Towards Research Platform as a Service (RPaaS) initiatives by Carol Gauthier and collaborators

76. Brandon University White Paper Submission to NDRIO

77. SOSCIP Consortium on the Importance of Driving Industry Innovation with Canada's New Digital Research Infrastructure by the SOSCIP Consortium

78. National Research Data Infrastructure in the context of Truth and Reconciliation by Ry Moran

79. Racism's trojan horse: race-based data, health care, and risks to Black and racialized populations by LLana James

80. On Canada's Future Digital Research Infrastructure by Ross Dickson

81. A Perspective on a Canadian Digital Research Infrastructure by the McGill Centre for Integrative Neuroscience McGill Centre for Integrative Neuroscience

82. Data Management Planning in Canada by James Doiron, Maggie Neilson, and Robyn Nicholson

83. Canadian Digital Research Infrastructure Needs Assessment: An Agri-food Research Perspective from the University of Guelph by Emily Duncan, Rozita Dara, Jairo Melo, Carly Huitema, and Beverley Hale

84. A Dynamic Environment for Medical Image Computing Research in Canada by Amber Simpson and collaborators

85. Building a data science platform for better health by Laura C. Rosella and collaborators

86. PHI-Compliant Computing and Storage: a critical need for Canadian biomedical and health research by Richard F. Wintle and colleagues for The Centre for Applied Genomics/The Hospital for Sick Children

87. Bilinguisme des outils et des services: soumission conjointe des Bibliothèques de l'Université Montréal et des Bibliothèques de l'Université de Toronto Bibliothèques de l'Université de Montréal et des Bibliothèques de l'Université de Toronto

88. ORION Digital research infrastructure for Canada by the ORION Network

89. People, Platforms, and Content: The Importance of Libraries in the Emerging National DRI by the University of Toronto Libraries

90. Persistent Identifiers in Canada – Position Paper by Eugene Barsky on behalf of the ORCID Canada Consortium, DataCite Canada Consortium, and the Canadian Research Data Management Community

91. DRI, University Libraries and Digital Humanities Research Centres by Jason Boyd on behalf of Ryerson Centre for Digital Humanities

92. White Paper on Canada's Future DRI Ecosystem: Subatomic Physics in Canada by Randall Sobie on behalf of the Institute of Particle and the Canadian Institute of Nuclear Physics

93. Successes and shortfalls of the current Canadian ARC platform and ideas to improve it further by Maxime Boissonneault

94. White Paper: Building and Supporting the Use of Digital Research Infrastructure Among Qualitative Researchers by Katherine Tamminen, Heidi Bohaker, Andrea Bundon, Denise Gastaldo, Brenda Gladstone, Cara Krmpotich, Meghan McDonough, Brett Smith

95. The Canadian Primary Care Information Network: Harnessing primary health care data to improve practice by the Canadian Primary care Information Network

96. The Canadian Research Data Centre Network (CRDCN) in Canada's DRI Ecosystem by Byron G Spencer on behalf of CRDCN

97. The Importance of Distributed ARC HQP by Chris Loken on behalf of Compute Ontario

98. The Need for First Nations Collaboration and Co-development in Digital Research Infrastructure (DRI) by the First Nations Information Governance Centre

99. Role of TRIUMF within the Digital Research Infrastructure Ecosystem by Jonathan A. Bagger, Reiner Kruecken, and Reda Tafirout on behalf of TRIUMF

100. Livre blanc sur l'infrastructure de recherche numérique : perspective de la Faculté des sciences de l'UQAM Faculté des sciences de l'UQAM White Paper McMaster Angela Di Nello, Research & High-Performance Computing Support (McMaster U)

101. NDRIO White Paper by Cape Breton University

102. Clés pour une Infrastructure de recherche numérique (IRN) innovante by Suzanne Talon, Marie-Jean Meurs, and Victor Ionescu on behalf of Calcul Québec

103. Les consortiums régionaux de bibliothèques universitaires sont des partenaires clés qui contribuent à la mise en place d'une infrastructure de recherche numérique durable et réactive by Vivian Stieda, Sonia Seck, and Cynthia Holt

104. A Cloud Ecosystem for Canada's Digital Research Infrastructure by Wency Lum and collaborators

105. DRI Needs Assessment for the Computational Fluid Dynamics (CFD) Research Community by the Computational Fluid Dynamics Society of Canada

106. NDRIO: Call for Position Papers on Canada's Future DRI Ecosystem by the Ontario Council of University Libraries

107. Advancing a Framework for National Commercial Cloud Adoption in Research by Jeff Gardner, Nick Rochlin, and Elizabeth Kinney

Appendix 3

Table of Contents

1 Introduction

Between May 4th and 7th the Alliance shared its preliminary findings from the National Survey and the position papers in a series of four Virtual Town Halls. These findings were organized in four themes: Awareness and Accessibility, Operations, Governance and Policy, and Support. All sessions were open to all members of the research community in Canada, regardless of their level of engagement with DRI. A total of 477 researchers, librarians, administrators, and representatives of funding agencies participated in these Town Halls.

Feedback during the Virtual Town Halls was gathered in breakout sessions where participants were able to contribute their perspectives, whether through commenting or directly in a shared document. The following is a summary of the main points gathered during the breakout sessions and open discussions. Given the extent of the feedback, comments, ideas and perspectives were organized thematically and consolidated into the edited notes below. Direct quotes are presented in italics. Together, these thoughts represent the collective effort of all attendees who provided their feedback.

2 Awareness and Accessibility

2.1 How to improve access to DRI

It is essential to have local liaisons who direct researchers to the appropriate national DRI resources available to them. Without the local layer of support, researchers would not be properly onboarded and walked-through the discovery of the tools and services the Alliance would offer.

Workgroups, training seminars, and related information sessions may have greater impact when coordinating with localized advanced research support groups. If researchers, including graduate and undergraduate students, are required to take ethics and safety courses, why not DRI?

> *"University of Toronto has courses for credit related to DRI. New university policies were needed to support the integration of DRI usage in training and throughout the institution. Additionally, University of Toronto set up resources locally, such as tools and guidance videos which were complimented with workshops provided across the university. Such model would be valuable if adopted across other universities."*

> *"Websites are additional ways to improve accessibility to DRI, but they usually become useful once researchers understand the Canadian DRI landscape. They serve an important purpose but are not the whole answer."*

Learning about tools and resources available to researchers is very important, but national platforms need to be interoperable and provide similar entry points and usability. A high-level approach to national platforms may help avoid duplication (platforms performing similar/same functions) and enhance interoperability.

DRI is increasingly focusing on data, but there is work to be done around networking (e.g., connecting data centers, download, upload data, workflows). Researchers use local resources because data download can be slower across a national network–affecting productivity, support, and access to national DRI infrastructure. Additionally, dependance on email or ticketing support poses important delays to researchers.

> *"I do not want to be a software developer or system administrator. We just want to do research and not wait for over 8 weeks for access."*

2.2 Bottlenecks in DRI-related funding

Stable and sustainable funding for computing resources for researchers is needed. Current allocations are too short (~ 18 months). In the Social Sciences and Humanities these limitations are especially important–both in the type of eligible expense and the amount in dollars. Moreover, it is currently unclear which DRI-related expenses are eligible for which funding agency. For example, the distinction between CFI and the Alliance is at the moment ambiguous.

There is a general feeling that funder expectations are often not feasible. For example, grants for infrastructure often require researchers to demonstrate scalability and to contribute to national compute clusters. There is, thus, an additional burden on researchers to demonstrate why they cannot follow such a model.

Funding for Research Software support and development is precarious in Canada, with few exemptions such as the CANARIE-funded Local Research Software Support program.

The Resource Allocation Competition process is cumbersome, slow, seems arbitrary, and takes tremendous amounts of human resources to complete. There is an opportunity to start anew and make the entire process more transparent, more local, and faster.

2.3 Channels for DRI discovery

Researchers discover the DRI tools and resources they use through word of mouth, or through recommendations and provision by their institution.

Research Offices in universities play an important role in helping researchers identify infrastructure opportunities during the development of grant applications. Given that communication through these Offices is often received and read by researchers within a given institution, this is an important channel to discover DRI tools and resources. Unfortunately, oftentimes researchers do not receive support from Research Offices outside grant-development. Some researchers then heavily rely on institutional IT or libraries services.

> *"There is a need to promote the integration and closer collaboration of IT, libraries, and the VP-Research Office in universities. Current service provision is very siloed in many institutions."*

There is a need to properly distinguish the roles and responsibilities between institutional IT and technical research support (i.e., ARC). At St. Francis Xavier University, for example, IT services are provided along with ARC services. Meanwhile, at the University of British Columbia (UBC), researchers receive help accessing both UBC's and national resources through their Advanced Research Computing services. Establishing similar offices and services in other universities would improve discovery and access to national DRI.

Regional consortia like ACENET play an important role in facilitating the discovery of DRI tools and resources, given that they work across institutions of different sizes and geographies.

3 Governance and Policy

3.1 Health data

In addition to long-term solutions to improve the sharing of sensitive and health data across provinces, there is the need for short term solutions that provide resources and governance that will fit within the current more restrictive environments of many of the health authorities. Alignment is a decade long process and researchers need to have solutions that can fit into the current policy landscape.

There is a need for infrastructure to securely host sensitive data, but also the ability to *access* it. Differences in legislation pose some barriers to using health related data, but in part it may be an overly restrictive interpretation of privacy policies.

There is need for a coordinated approach across a variety of sectors to ensure that data can be accessed.

Researchers need specialized guidance and support from professionals with judicial understanding of the differences between the different legislations applied to sensitive data.

The Advanced Research Computing needs are not being met for health researchers. Specifically, Compute Canada resources cannot be used for sensitive data that require restrictions. As a result, many research teams build small *“mom and pop”* shops for RDM and ARC.

Many researchers use micro-level data from Statistics Canada through the Canadian Research Data Center Network (CRDCN) and their Research Data Centers (RDCs). Computing power in RDCs is limited, statistical packages or often outdated or cannot run machine learning algorithms or modelling. This prevents researchers from using the latest methodologies and computing infrastructure to analyze the data. Moreover, RDCs require physical access to their premises to perform analyses, which under current sanitary restrictions and lockdowns has been extremely difficult to achieve.

There is need to make it possible for health researchers to access existing and shared resources.

Otherwise, things are done in a piecemeal fashion–which can affect interoperability.

A federated approach for the analysis of health data–that is rather than taking data to where analysis is done, leave the data where they begin–would be a valuable adoption that may circumvent some jurisdictional issues. However, whether this can be done will have a lot to do with the architecture set up and supported.

> *“The Alliance could play a role in establishing practices and policies for supporting distributed architecture, and also play a role in setting it up.”*

Consent is an issue often overlooked. Participants may not have consented to allowing their data to be used for all purposes. Consent should be built into the process as it is difficult to do this retrospectively. Must be front of mind in system design.

3.2 Indigenous data sovereignty

The governance system needs to understand, learn, and establish a relationship with First Nations, Inuit, Métis, and other Indigenous Peoples. There is need to continually keep the contact and build meaningful relationships.

First Nations, Inuit, Métis, and other Indigenous Peoples are not research subjects and should not be considered so. They are the ones to help and lead the research. The Alliance ought to adopt a model of engagement of “*no research about us without us”.*

Relationship building has to come first, researcher to community level, and researcher to infrastructure. Indigenous Peoples need to be engaged and actively participate in the development of DRI plans, to make sure that their governance systems are considered and implement, and that their rights are not infringed or superseded. Furthermore, the Alliance should be there to help nurture those relationships and lend a guiding hand, but those relationships need to be able to flourish. *"There is a nurturing aspect which happens behind the scenes, and that would be a success."*

There are examples of good collaborations to support research, especially Indigenous-led research. There is substantial federal funding recently awarded towards Indigenous data sovereignty, and Indigenous groups will be developing their own data strategies. This is an important opportunity for the Alliance to collaborate with these groups to ensure they are not left out of their strategic planning. This work cannot be done in a centralized way; it needs to be distributive, well supported, and within the Truth and Reconciliation effort.

FNIGC has played a role in rethinking how First Nations data ought to be governed, given that traditional consent models may be inappropriate. First Nations have taken a key leadership role in this. Community consent and individual consent each have a role in data governance.

Indigenous Peoples have distinct needs and there should be an allowance for nuance. Tiered access, for example, provides some of this nuance and flexibility.

Traditionally, in the DRI space, the Compute Canada Federation and its regional organizations have focused on technical solutions and the digital research infrastructure only. There are important gaps in the policy aspect– *"the best technology won't amount to anything important if the relationships and data sharing agreements are not in place*".

It would be meaningful to enact or support enactment of technology and other services such that Indigenous Peoples can lead the work on data sovereignty. There has been a lot of talk over time about commitment to data sovereignty and supporting Indigenous data sovereignty, but this has not been backed up with the resources that would be necessary to enact this in practice.

3.3 Alignment of governance policies, practices, and procedures.

The Tri-Council Policy on Research Data Management is creating a two-tier system of researchers as it only applies to Tri-Agency funding. It limits the amount of control that universities have to regulate research. A policy to encompass all researchers is needed.

One important problem is that the ownership of the data is not clear to researchers. In Compute Canada, for instance, there is the understanding that each hosting site and institution have their own policy. Compute Canada does not currently have a clear policy to change data ownership without having problems with Intellectual Propriety, if for example the original owner cannot be contacted anymore (e.g., was not responding to email or might have moved to different institution). There is need for a *"next of kin for data since we had sadly PI and user passing away and have no idea what to do with their data".*

Compute Canada is currently drafting data classification policy but there are several questions and issues related to the type of data currently being hosted at their clusters. Data ownership and

Intellectual Property should be considered for a unified policy for all different data sets, in different regulated areas and across host sites.

Building a risk management framework and learning decision making tools could allow for a much more open and free data discussion.

A national-level legislation could be adapted, and a discussion around how the Alliance could play a great role in building a platform that would automatically generate legitimate stamped certificates that would then be used by federal agencies to expedite access processes.

3.4 Public-Private collaborations

Publicly funded research using commercial cloud or conducted in collaboration with industry should not have a detrimental effect on academic researchers. There is a limit to the overall number of resources in the private and academic clouds, and so any public-private relationships should bring resources into the system.

Many commercial partners do not need any additional computing infrastructure, since they already have access to faster and more powerful systems. What these partners are often missing are the Highly Qualified Personnel (HQP) and Professional Support Staff.

In order to drive industry-academic/post-secondary adoption of Artificial Intelligence and data sciences and spur manifold commercial, economic and societal benefits, the Alliance should take leadership and provide long term stable support for a pan-Canadian industry-academic DRI operating model based upon the successful model built in Ontario over the past decade by the SOSCIP Consortium. In this model, industry-academic collaborations should be supported through dedicated priority access to their own separate ARC DRI resources, which would be entirely complementary to discovery-based ARC research support. It would also enhance further industry engagement and deepen ties with universities and colleges including the training and hiring of HQP and Professional Support Staff. Company and partner accessibility to reliable DRI is critical along with the security of the data that would reside on it.

> *"Requests for higher levels of data security have become common at SOSCIP in recent years, not only for health-related projects but from virtually every sector where we have worked. Recent cyberattacks have further increased demand for cybersecure ARC DRI to the point that clients are generally inquiring about it for virtually all collaborative project work."*

From an international perspective, it is hard to collaborate across borders when one country has invested in cloud services and infrastructure, and the other has not. Canada should lead by example and expand its cloud services provision.

Systems must be built with clear porting mechanisms in place to avoid vendor lock-in.

Current access to national infrastructure is restricted to academic researchers in academic institutions. Consideration should be placed in improving access to academic researchers outside such sector (e.g., non-profit organizations).

3.5 Adoption of Open Scholarship

In the context of Open Scholarship, data sharing is not the end-product. Open methods, open calls, and overall transparency are also essential. Providing a plan or a template could incentivize and help people do this type of work. Tenure promotion can be a major incentive.

> *"The lack of recognition and approval in this kind of work has become a real problem stopping researchers in their process. It should be taken care of at an institutional level, and they are the ones that should answer the concerns"*

Medium and small data are important too–not just big data.

The Alliance should adopt the FAIR and EDI principles and require their implementation in all funded projects.

4 Operations

4.1 Collaboration arenas

There needs to be a distinction between the IT support and Research Support. There also needs to be closer collaboration between these two sectors, so that researchers are directed to the appropriate personnel.

Researchers also need a direct channel for communications with the Alliance, since there are many researchers at small institutions that do not have the resources to have effective intermediary communication.

Working groups with an end date and a specific objective are excellent mechanisms to tackle specific issues.

Webinars can be informative but are impersonal. It is hard to engage with panels.

The Alliance should develop clear mechanisms for feedback and communication between the broader research community and the Alliance's Researcher Council.

4.2 Key Performance Indicators (KPIs)

The broad scope of the Alliance's activities is a challenge. The Alliance should do a set of things well, and not try to do everything for everyone since from an operational perspective, this is probably not feasible.

There must be equity amongst the various portfolios and research disciplines.

Consider disciplinary differences–the measures of excellence are different for various fields.

While developing KPIs, the Alliance should not just benchmark against past performance but also against international organizations with similar mandates.

Researchers in the Humanities cannot compete with Sciences for resources if the criteria for excellence are size of data outputs, number of research outputs or even use of computational power. Humanities are putting out a different number and size of outputs.

Develop KPIs derived from researchers' perspective, and not from the organization itself.

4.3 Transparency

Considering carbon emissions of computing resources.

There is very little to no input into who is using national computing infrastructure–it is all decided by the RAC committee.

The Alliance should be the one reaching out to researchers for feedback and understanding their needs. It should not be the researchers reaching out to the Alliance.

It's important to realize the Alliance is not Compute Canada 2.0. There are more activities in the Alliance's portfolio, such as RDM and RS.

> *"Not obvious to all where the additional budget is coming from, and who will decide where the money goes. Researchers want to see how decisions are made into investing into the priorities, and who is making those decisions. Communicating clearly - what are the short/near/long term priorities? How are the board and executive deciding what to prioritize?"*

> *"A lot of people are hoping to get efficiency and standardization across different groups. Because the Alliance is publicly funded, we don't want to see siloed investments by the federal government that are doing similar things. Need transparency on how funds are used."*

The Alliance should outreach and provide incentives for institutions to participate in the national DRI. Absence of support staff should not preclude smaller institutions from using national resources.

Not all institutions have appropriate resources and staff to support HPC. Institutions may be willing to collaborate with others to increase the services they are able to provide.

Publicly facing interactive dashboards that show use by user and topic characteristics could provide researchers and public with a glimpse into resources allocation and increase trust and accountability.

5 Support

5.1 Collaborative network of support

For training and collaboration, it would be fruitful for closer collaboration between Portage and Compute Canada communities together.

Annual training in the Digital Humanities (the Digital Humanities Summer Institution) has been providing peer-to-peer training and has engaged in collaborations with Compute Canada. This model could be replicated and scaled up across disciplines.

Researchers often consult with Compute Canada staff on computing challenges. This model puts more burden on support personnel who have to consult with each individual researcher, rather than to train researchers on these areas themselves.

Faculty members often don not have time to engage in as many resources and training opportunities. Graduate students are crucial in making these connections for faculty.

A “research month” at the beginning of the year, supported by the Vice-Principals Research, aimed at graduate students and early career researchers could be an important introduction to national infrastructure and best practices. This may be a part of a credit course for graduate students.

The Portage- Alliance National Training Expert Group materials are widely available. But it is hard to figure out what priorities are.

Software carpentry courses also teach researchers the basics and necessary skills.

There is a need for overlap of Research Software and Project Management to assist researchers in setting up such frameworks. Local Research Software support is making a great difference with continuous progress as they are handling the front end of the data pipeline.

If highly qualified personnel rely on grants, they will always have issues with retaining qualified personnel because they are looking for stability - and grants don’t provide that. Need buy-in from the top of the organization for funding vs. grant to avoid worrying if their role is going to be renewed via grants - you’re more likely to see talent leave for the private stable sector.

Academia cannot compete with private sector jobs when it comes to salary expectations. This contributes to turnover, but also it isn't fair to those who are staying due to passion projects but are earning a lower salary.

5.2 DRI Champions

This should not be a volunteer proposition. They need to be taken seriously and feel supported. The Alliance could hire 10 champions and deploy them across the nation, or the funding could be distributed to institutions to hire data champions. From the administrative perspective of the institution, there is institutional responsibility to ensure and safeguard these roles. Financial support for people who engage in technical support activities could be another option.

It is very important to have institutional representation.

"Champions" may not be the best term. Consider using ambassador or similar term instead. Can also extend the idea beyond 'data' or DRI. For example, have specific ambassadors for software, code, or compute.

These roles need to have clear and strong connections between local and national support. They need to understand what services are best provided locally *vs*. nationally.

DRI Champions could be regional and national experts that are organized and offer training, presentations, and consultation services. Rationalize national and local training opportunities to share resources, best practices, and to avoid redundancy.

Develop a community of practice composed of a mix of researchers and research support personnel in RDM, ARC, and RS. The Alliance could play a convenor or coordinating role for these communities of practice.

Champions should be thought of in terms of networks of champions.