Developing a digital research infrastructure strategy for Canada: The CFI perspective

November 2015

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

Table of Contents

LIST OF ACRONYMS

CARL	Canadian Association of Research Libraries
CERN	Conseil européen pour la recherche nucléaire (European Organization for Nuclear Research)
CFI	Canada Foundation for Innovation
CIHR	Canadian Institutes of Health Research
CISTI	Canada Institute for Scientific and Technical Information
CUCCIO	Canadian University Council of Chief Information Officers
DRI	Digital Research Infrastructure
IC	Industry Canada
MSI	Major Science Initiatives Fund
NEPTUNE	North-East Pacific Time-Series Undersea Networked Experiments
NRC	National Research Council of Canada
NSERC	Natural Sciences and Engineering Research Council
ORANS	Optical Regional Advanced Networks
RDC	Research Data Canada
SPARC	Sustainable Planning for Advanced Research Computing
SSHRC	Social Sciences and Humanities Research Council

ABOUT THE CANADA FOUNDATION FOR INNOVATION

The Canada Foundation for Innovation gives researchers the tools they need to think big and innovate. By investing in state-of-the-art facilities and equipment in Canada's universities, colleges, research hospitals and non-profit research institutions, the CFI is helping to attract and retain the world's top talent, to train the next generation of researchers, to support private-sector innovation and to create high-quality jobs that strengthen the economy and improve the quality of life for all Canadians.

EXECUTIVE SUMMARY

Researchers, ranging from astronomers and sub-atomic physicists to economists and linguists, are increasingly reliant on the production, sharing and management of large amounts of data to generate knowledge. As a result, there is a growing need in Canada to ensure our brightest minds have access to an efficient and effective digital research infrastructure ecosystem. The Government of Canada is addressing this need by developing a digital research infrastructure (DRI) strategy that will enhance the country's ability to employ digital resources across all areas of research and will advance the Digital Canada 150 plan.

Over 20 years, Canada has developed a DRI ecosystem that has made notable advances, such as developing a world-class national high-speed backbone network and offering researchers access to a range of advanced research computing resources. But the accelerating pace of both technological change and evolving needs of the research community has led to the development of a DRI ecosystem, involving numerous actors who share overlapping roles and mandates, that is not yet functioning optimally.

To ensure a high-performing DRI ecosystem in Canada, the CFI envisions an integrated system that promotes greater coordination and alignment between its individual components and actors. This can be achieved by adopting a holistic, long-term focus on promoting the DRI ecosystem capabilities.

From the CFI's perspective, there are seven essential attributes of a national high-performing DRI ecosystem: integrated, inclusive, sustainable, comprehensive, accessible, user-centric and adaptable. It should also respond to the needs and priorities, existing and emerging, of the Canadian research community and have robust governance and management, and predictable multi-year funding. This will allow for greater coordination and alignment among the components and actors and will ensure that users have timely access to the appropriate resources and services.

The CFI also recommends concrete, strategic objectives and short- and medium-term actions to better align and coordinate the actions of all actors in the ecosystem, and to ultimately enhance our country's collective capability to conduct computationally-challenging and data-intensive research.

THE CONTEXT

A changing research landscape

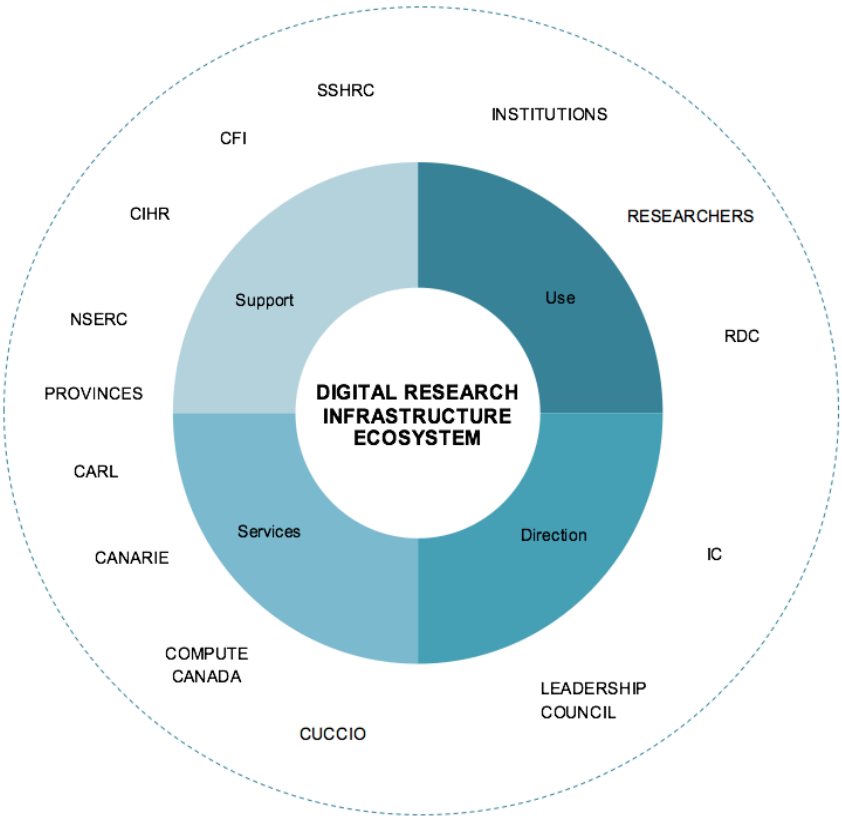
Today, conducting leading-edge research increasingly requires the use of large-scale, shared research facilities and the management of large amounts of data. In many fields, a growing number of researchers are coming together in large groups or consortia to produce, share and manage their research data. To draw insights and advance knowledge, they now rely on the availability and robustness of research data sets, data platforms and analytical tools. To address this evolution in the research environment, the Government of Canada has called for the development of a digital research infrastructure strategy that will enhance the country's ability to employ digital resources across all areas of research and will advance the Digital Canada 150 plan.

What is digital research infrastructure?

For the purpose of this paper, the CFI defines digital research infrastructure (DRI) as the components that are collectively managed and operated as shared facilities and services for research institutions and users across the country because they are so large in scale, complexity and cost that they cannot be offered by a single institution. These components comprise computational capacity, data storage, technical services, research software, middleware, high-speed optical networks and research data management capabilities. Computing resources and assets devoted to the work of individual researchers and small teams with modest requirements fall outside of this definition since the CFI considers these to be the responsibility of individual institutions.

Canada's DRI: The assets and actors

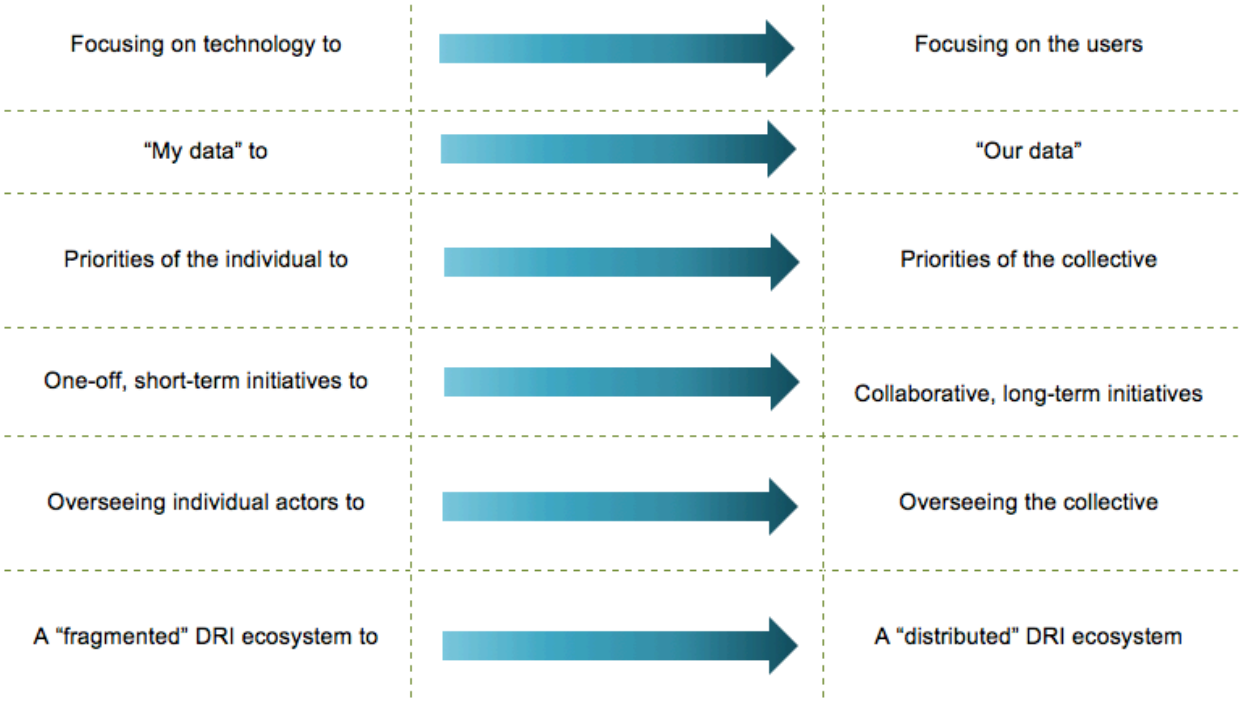
Over the past two decades, Canada has acquired significant assets and developed advanced capabilities for a number of components of DRI: a high-speed backbone network, advanced research computing infrastructure, large collections of valuable research data, analytical technologies and tools, and the creation and enhancement of several independent stakeholder organizations. Canada's complex DRI ecosystem involves a diversity of stakeholders, or "actors," each with different mandates and roles (see **Appendix 1** for an overview of the roles and responsibilities of the current actors).



The challenge

The transformation in the growth and evolution of Canada's digital research infrastructure ecosystem is, in some cases, leading to true paradigm shifts in how research is being planned, conducted and measured. To meet the demands of an evolving DRI ecosystem, these shifts must be carefully considered.

Shifts in the digital research infrastructure ecosystem



In addition, the advent of this era of "big data" and the prevalence of data-intensive and computationally-intensive research has led to increased attention and support for research data infrastructure, software tools and data management, along with the development of the required skilled personnel to exploit these capabilities. The principal challenge lies in the fact that the many actors who share these responsibilities have overlapping roles and mandates. While the current approach to DRI in Canada has allowed different actors to operate and evolve independently, the strong performance of these individual actors has not translated to success for the DRI ecosystem as a whole. This independent approach has been regarded by many in the research community as the underlying reason for a "fragmented" and "disorganized" ecosystem that "lacks cohesion."

THE STATE OF CANADA'S DRI ECOSYSTEM

In many respects, the Government of Canada's approach for DRI has been successful. Through a variety of funding and policy mechanisms, it has supported the development of a range of DRI resources and assets. For example, 20 years ago, oceanographic observatories didn't exist. Since then, Canada alone has pioneered two — the Ocean Tracking Network and the Ocean Networks Canada. Over the same time period, Genome Canada was created, and now three world-class regional genomics centres produce huge amounts of data every year. And Canada hosts one of the Tier 1 ATLAS GRID computing centres to

manage and exploit the petabytes of data generated each day by the supercollider at the European Organization for Nuclear Research.

The funding approach has also led to the development and operation of a world-class national high-speed backbone network, as well as shared national advanced research computing resources that all researchers, irrespective of their institutional affiliations, can access.

Today, Canadian researchers have access to seven DRI components:

1. A national ultra-high-speed backbone network operated and managed by CANARIE and its provincial and territorial Optical Regional Advanced Networks (ORANS) which support data-intensive, leading-edge research across Canada;
2. Computational and data storage resources and support services overseen and coordinated by Compute Canada, which include management and technical support, irrespective of their institutional affiliation;
3. Resources and tools to build shared research data infrastructures;
4. Research software and software analytics to enable users to process and analyze research data;
5. Middleware to manage, share and move research data across the country and internationally;
6. Research data management policies, practices, methods and supports; and,
7. Highly skilled and qualified personnel to support all aspects of digital research infrastructure.

The first two components — the network and the computational resources — represent the most significant assets of the DRI ecosystem, and accordingly, have received the majority of government support and attention.

High-speed network: CANARIE was established by the Government of Canada in 1993 to operate Canada's advanced network, working closely with provincial and territorial network partners to enable data-intensive research across the country. Through predictable multi-year federal funding over the last two decades, Canada has built a world-class high-speed network.

Advanced research computing: Since the early 2000s, the CFI has been the primary source of funding for large-scale computational resources. Between 1999 and 2006, the CFI's open research infrastructure competitions allowed institutions to acquire high-performance computing systems. But recognizing that greater coordination and integration would optimize the use and access to these large-scale systems, the CFI held a special National Platforms competition in 2006, and provided $78 million for an integrated advanced research computing infrastructure platform. This investment led to the creation of Compute Canada.

The world-class pan-Canadian advanced research computing platform that resulted from this investment was limited in its effectiveness by the absence of predictable multi-year funding to support infrastructure renewal and modernization. Recognizing a growing concern, the federal government announced $50 million of targeted funding in 2013 and an additional $100 million in 2015 for advanced research computing through the CFI's Cyberinfrastructure Initiative. In June 2015, the CFI invested $30 million of this amount to address the most pressing infrastructure needs of the pan-Canadian advanced research computing platform. (In addition to this support, the platform received funding in 2010 through the CFI's Major Science Initiatives Fund competition, which provides predictable and multi-year operational funding for national research facilities.)

In developing its Cyberinfrastructure Initiative, the CFI focused on the collective needs of communities of researchers in building and exploiting massive volumes of research data — often referred to as the "data deluge." By challenging institutions and researchers to form consortia, and to propose projects that make use of tailored, shared and integrated research data infrastructure, the CFI adopted a bottom-up

approach driven by research collaboration. It is shifting researchers' mindsets from a "my data" to an "our data" perspective.

While there has been some success on the DRI front in Canada, challenges remain which necessitate the development of a strong DRI strategy. Besides gaining access to CANARIE and Compute Canada resources, the research community needs to focus on the remaining five components, which are just as critical in ensuring a high-performing DRI ecosystem.

Resources and tools to build shared DRI: The capacity of researchers to generate data, individually and collectively, has grown exponentially over the past decade. They have access to a world-class high-speed network that has benefitted from predictable long term funding, but the advanced research computing resources available to them are no longer state-of-the-art. This is, in part, because the pace of investment in research computing has not kept pace with the rapidly decreasing useful life (typically 3 to 5 years) of computing infrastructure. As a result, many researchers have had to scale back and tailor their research to the computing resources available, limiting their ability to exploit the potential of large data sets.

Software and middleware: Investments in research software and data management software tools are critical to fully exploit the potential of large data sets and make the most efficient use of Canada's backbone network. Previous investments by the CFI and CANARIE, and more recently by Compute Canada, have undeniably facilitated the access, sharing and movement of data both nationally and internationally. However, these investments have only partially met the demand for software tool development. Meeting these demands is not simply a matter of increasing support for the CFI, CANARIE or Compute Canada.

Take the recent CFI Cyberinfrastructure Initiative competition: a third of the proposals invited by the CFI build on software tools and data platforms previously funded by both the CFI and CANARIE. The CANARIE investments allowed targeted research software development efforts to support and enhance significant CFI investments made in the early 2000s, such as McGill University's CBRAIN and the University of Victoria's NEPTUNE ocean observatory. In effect, the CFI provided the foundational infrastructure and CANARIE extended their software capabilities. Now, the CFI is scaling up research data infrastructure and software tools through its new cyberinfrastructure competition. This sequence of events, which will undoubtedly lead to excellent outcomes, occurred serendipitously, in absence of a DRI strategy.

Research data management: As arguably the most important of the seven DRI components, research data management remains an area that has not received the attention necessary to achieve a high-performing DRI ecosystem. Since every actor in the DRI ecosystem has some level of responsibility and mandate for research data management, there has been fragmentation in this area. The Research Data Canada (RDC) was created in an attempt to address this problem, but it has enjoyed very limited success. Recognizing this gap, CANARIE recently stepped up and committed to creating the position of Executive Director to enhance RDC's efforts and profile. But is CANARIE best positioned to take on this role? It could be argued that the three federal granting agencies have a vested interest in contributing financially to RDC's success and ensuring that the organization is properly mandated, managed and governed, but in the absence of alternatives, CANARIE acted to fill a void.

Support of highly qualified personnel: The growing demand for highly qualified personnel in this field is the result of the convergence of several trends, including the increasing number of researchers working in predominantly digital environments, advances in scientific and biomedical instrumentation and the declining cost of computing and storage. These trends impact areas of research ranging from those that use remote sensing and sensor networks to virtual engineering systems and digital imaging systems. These trends also explain the growing demand for analysts with discipline-specific knowledge who can assist researchers in making appropriate choices for advanced computing technologies and advise them

on how to customize advanced computing software and who can also provide training to meet the needs of research teams.

THE CFI VISION FOR DRI

To move national digital research forward, the CFI proposes that Canada's valuable digital research infrastructure assets and research capabilities be managed as an integrated system that promotes greater coordination and alignment between its individual components and actors. This can be achieved by adopting a holistic, long-term focus on promoting the DRI ecosystem capabilities.

The CFI considers that there are seven essential attributes of a national high-performing DRI ecosystem: it should be integrated, inclusive, sustainable, comprehensive, accessible, user-centric and adaptable[1].It should also respond to the needs and priorities, existing and emerging, of the Canadian research community and have robust governance and management, and predictable multi-year funding. This will allow for greater coordination and alignment among the components and actors and will ensure that users have timely access to the appropriate resources and services.

A holistic, ecosystem approach for the development and management of DRI should consider the components and actors, the roles they play, the interactions among them and how these interactions affect the overall performance and health of the system. Improving the coordination and alignment of activities and investments will ensure the ecosystem can optimize:

- Effectiveness: the ability of the actors in the ecosystem to carry out their respective mandate and deliver their expected results.
- Efficiency: the actions and decisions that maximize the performance of the ecosystem as a whole by reducing duplication and overlap, and enhancing overall coordination and alignment.
- Equilibrium: the actions and decisions that take into account interdependencies to ensure the entire ecosystem is in balance.

AN INTEGRATED, COORDINATED DRI ECOSYSTEM

DRI evolves quickly. Canada has many of the components in place to respond effectively to this evolution. The next natural step is to better integrate and coordinate the various components in the ecosystem to fully realize the benefits from the significant DRI investments already made. The CFI is convinced that a high-performing DRI ecosystem in Canada should be both informed by and structured around community-led research planning exercises that address the "data deluge" challenge. It must also address the pressing need for the development of highly qualified personnel.

Informed by user needs

In the era of big data, researchers approach the process of discovery differently than they used to. More and more researchers produce, share and manage their research data in large groups and consortia that depend on the availability and robustness of widely used data sets, data platforms and analytical tools to draw insights and knowledge. The research world in which astronomers and particle physicists have operated for decades, for example — where large data sets are shared and used by a large group of

…….....

[1] See Appendix 2 for a description of the seven attributes of a high-performing DRI ecosystem.

researchers — is becoming commonplace for researchers in health and the sciences to engineering and social sciences. This trend underlines the need for collective research infrastructure development and data management planning.

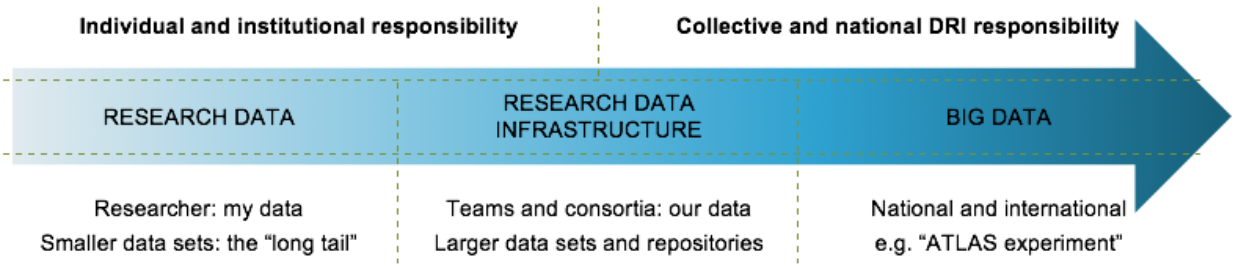
Research planning exercises are a critical component of a high-performing digital research infrastructure ecosystem. Yet, there are few incentives currently in place for research communities to engage in research planning exercises that identify areas where, collectively, Canadian researchers have a depth and breadth of expertise to strengthen our competitive advantages. Adopting such community research planning exercises would help structure and focus the DRI ecosystem by identifying priorities and requirements for networking, computational and storage resources, user support and highly qualified personnel. Such planning would also enable research communities to determine collective research data infrastructure and data management needs. This approach would be similar to the successful planning exercise led by the CFI in January 2014 on the future development of cyberinfrastructure for research in Canada. A similar planning exercise was undertaken by Compute Canada in 2014 in an attempt to assess future needs of specific research disciplines; the Sustainable Planning for Advanced Research Computing (SPARC) could provide the basis for such research planning exercises.

Planning for and managing the data deluge

Canada needs a sound DRI strategy to manage the unprecedented volumes of research data being produced. Our technical capability to generate data on a vast scale reduces the unit cost of data, but it creates data pollution along with the data deluge **—** we may have the ability to produce vast amounts of data, but that should not mean we need to store and preserve it all. It is only through proper research planning, data management policies and incentives that we can ensure the data generated is used and preserved appropriately.

The production of research data should be viewed as a continuum, from small-scale data managed by individual researchers to big data that can only be managed through multinational collaborations, as in the case of the ATLAS project at CERN. Institutions continue to be responsible for supporting their researchers' individual needs at the smaller end of the scale — commonly referred to as the "long tail of data" — while Canada's DRI has focused on the other end of the scale.

Management of research data



Smaller research data sets produced by individual researchers and small teams at the single lab level often have value for the collective research community, but it remains untapped. It is the responsibility of the researchers producing that data, along with their institutions, to provide access to their data and

metadata. However, without agreed-upon standards and terminology, these data cannot be aggregated into a collective resource and are of little value as a shared resource.

Advances in the development of data management standards and practices are critical to a high-performing DRI ecosystem. A first modest step towards addressing that question is the policy on research data management developed by the three federal funding agencies (the Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council and the Social Sciences and Humanities Research Council), which stipulates that the proper care of individual data sets is the responsibility of researchers and their institutions. The bottom-up approach of the CFI's Cyberinfrastructure Initiative provides an opportunity for consortia of researchers in specific domains to collectively develop such standards and methodologies.

At the same time, there have been repeated calls over the past 15 years to develop national data preservation and archive infrastructure[2]. Although preservation and archiving are important elements to consider, they fall outside the CFI's definition of research infrastructure. As a result, these are excluded from the CFI vision for a national DRI strategy.

It's about people: Addressing the growing demand for highly qualified personnel

In the big data era, data experts are invaluable. This is perhaps the biggest aspect of the research data management challenge we face. Better coordinating and integrating DRI in Canada will help address the need to support the highly qualified personnel who optimize the productivity of the system. The human resource dimension of DRI is as important as the infrastructure itself.

Generally, researchers are not, nor should become, technical experts in data and software. They require support from applications analysts to properly configure their data, and from experts to develop the software tools to access and fully use the data. The need for such support is even greater in fields that have not typically been data-intensive. Ready access to highly qualified personnel would allow researchers to focus on research rather than spending time mastering advanced computing and data management techniques.

Exploiting the full potential of advanced research computing systems has become more challenging with the use of more web-based applications that rely on sophisticated data analysis and visualisation tools. Beyond the highly qualified people needed to operate and maintain computing centres, today's advanced research computing requires data analysts, data management experts, bioinformaticians, computational physicists, parallel programmers and others. This expertise is in short supply around the world. A national DRI strategy should help ensure that this critical resource is addressed by taking stock of existing resources and training support, finding ways to increase the number of experts and creating stronger links between researchers and the specialists they need to maximize the use of the country's advanced research capabilities.

WHAT IS INVOLVED IN A "DISTRIBUTED" DRI ECOSYSTEM?

In the distributed DRI ecosystem the CFI envisions, all service providers and funders would remain independent. They would, however, be assessed in the context of the DRI strategy's overarching objectives and priorities as part of a comprehensive periodic review of the ecosystem's overall

……..

[2] SSHRC, May 2001: National Research Data Archive Consultation, Phase One: Needs Assessment Report

performance. As a good practice, the results of this assessment would be made public. Similarly, needs and requirements identified through community-led research planning exercises would be compiled, assessed and prioritized. While specific stakeholder groups and research constituencies will advocate for their own needs and priorities, a high-performing ecosystem would have a comprehensive and system-wide framework to assess future needs. Monitoring the ecosystem's overall health and performance is a critical oversight function, providing a mechanism to quickly address emerging issues and imbalances.

Who should assume responsibility for comprehensive oversight? This responsibility could rest with the Science and Innovation Sector of Industry Canada. The granting agencies (with the exception of CIHR), the CFI and CANARIE are all currently under its current purview. The Sector has also taken on the responsibility of developing a DRI strategy as part of the Government of Canada's 2014 Science, Technology and Innovation Strategy. It would, therefore, be entirely appropriate for the Sector to be responsible for assessing the extent to which the ecosystem as a whole is achieving its objectives and addressing its priorities.

Long-term planning and funding horizon

A high-performing DRI ecosystem requires a long-term horizon for planning and funding of at least five to seven years to allow priorities, such as coordinated software tool development, human infrastructure and shared research data infrastructure to be optimally addressed. This long-term perspective lends stability and predictability to the system and is conducive to greater stakeholder engagement. Research communities are more likely to engage in planning activities, for example, if the strategy allows for a long-term view and is committed to a user-centric, science-based approach with sustained support.

A long-term view also has funding advantages; it would provide an opportunity to have actors, such as Compute Canada and CANARIE, on a coordinated five- to seven-year investment plan which would include periodic performance reviews. Periodic reviews offer an opportunity to measure progress in addressing the strategy's goals and priorities and allows for timely action on evolving needs and issues, which emerge as a result of scientific developments, technological advances and political considerations.

MAKING IT HAPPEN: A FEW RECOMMENDATIONS

The current approach for DRI in Canada aims to promote innovation and nimbleness as its principal advantages. A national DRI strategy that defines key objectives and priorities for the ecosystem will serve to strengthen these advantages, as will the adoption of a long-term perspective. But this perspective, alone, will not be sufficient to ensure a high-performing DRI ecosystem. The roles and activities of each actor in the ecosystem should be well-defined and their performance assessed, individually and collectively, against the objectives and priorities of the strategy. This should also include their willingness and ability to engage in collaborative and concerted activities.

Objectives and actions

In order to better align and coordinate the actions of all actors in the ecosystem, the CFI suggests the following concrete, strategic objectives and short- and medium-term actions:

Objectives:

- Foster an ecosystem that is user-centric and informed by community-led research planning exercises that define priorities and requirements for DRI in Canada;
- Ensure that key components of the ecosystem are supported in a coordinated, predictable, balanced and sustainable fashion; and,

- Provide comprehensive oversight, including public reporting, on how well the ecosystem is functioning.

Short- and medium-term actions:

- Define the mandates, roles and responsibilities of each of the digital research infrastructure actors, focusing first on service providers and funders;
- Adopt and implement a predictable multi-year investment plan for computational, data storage and network infrastructure, and their operation and maintenance, all the while ensuring that they are aligned with the evolving needs of users;
- Support community-led research planning activities designed to define functional requirements and priorities for both computationally-intensive and data-intensive research;
- Determine the appropriate balance between the network and the computational resources by consulting both users and service providers;
- Develop and support shared research data infrastructure in areas where Canada has expertise;
- Forge a coordinated approach to the development of research software and middleware, in conjunction with Compute Canada, CANARIE, funding agencies and user communities;
- Identify requirements and means that will enable the federal granting agencies to successfully implement their policy on research data management (planning, access and stewardship); and,
- Incent service providers to help strengthen the human resources behind Canada's digital research infrastructure by expanding their training activities and resources for job placements and internships.

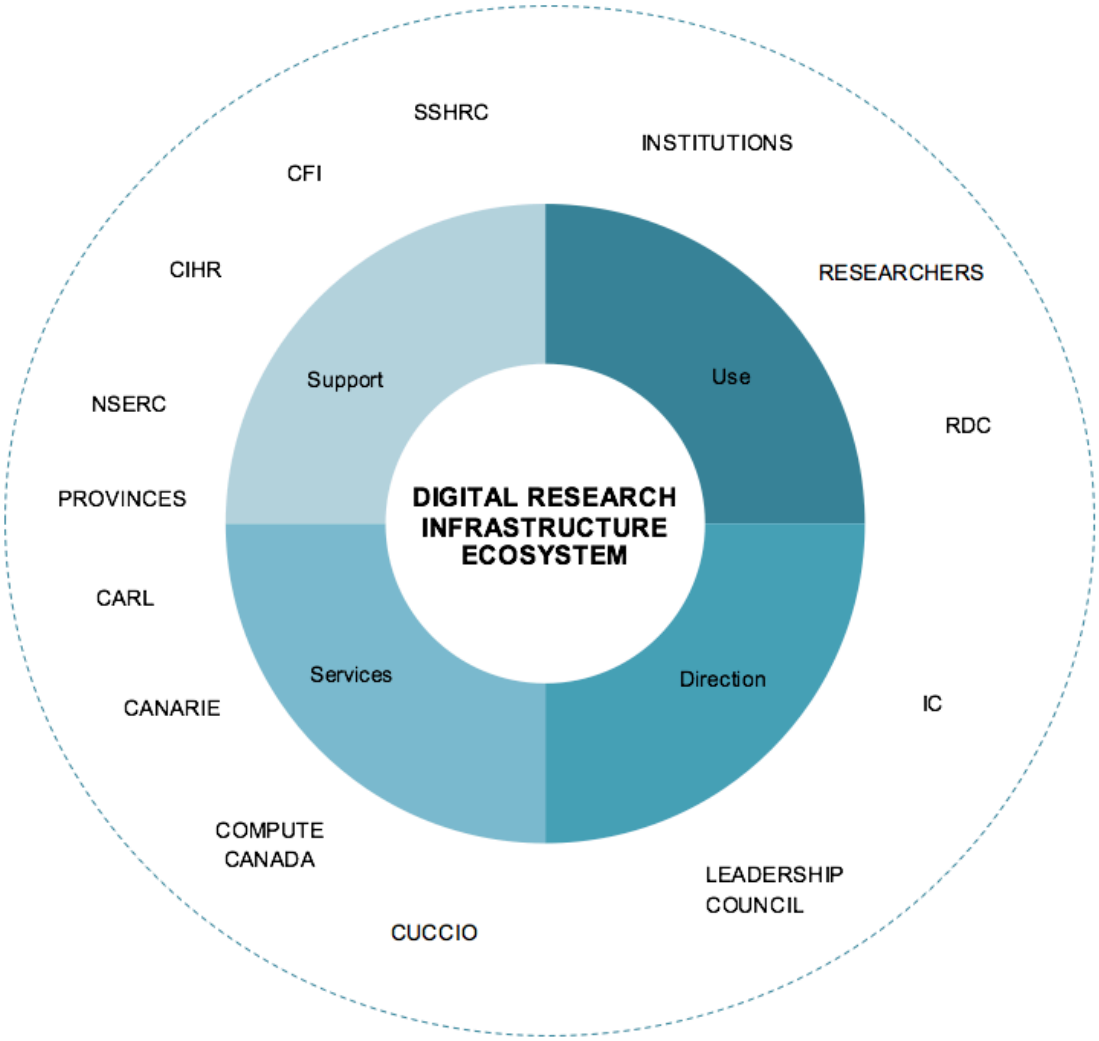
CONCLUSION

Canada should be proud of the achievements and outcomes it has made from its various DRI initiatives. With a world-class national ultra-high speed backbone network, access to a wide-range of advanced research computing resources, and financial support in place for the technical, operational and management services required to exploit DRI capabilities, the system has served its users adequately. The system is now sufficiently mature for its actors to start leveraging the strengths and specific contributions from one another. A DRI ecosystem that is integrated, inclusive, sustainable, comprehensive, accessible, user-centric and adaptable would enhance our country's collective capability to conduct computationally and data-intensive research.

APPENDIX 1: ACTORS IN CANADA'S DRI ECOSYSTEM

There are six major stakeholder groups in DRI in Canada: researchers and users, institutions, service providers, allied organizations and associations, funders, and provinces and regional economic development agencies of the federal government.

Actors in the DRI ecosystem



Researchers and users

Researchers and users from all disciplines are the primary stakeholders of the country's digital research infrastructure; the system exists as a foundation that supports their research activities. This group is characterized by its diversity. It includes astronomers, cosmologists, sub-atomic physicists, high energy physicists, bio-informaticians, computational biologists, genomics researchers, engineers, economists, historians, linguists, and many others.

When it comes to DRI, however, each of these user communities has very different needs. Some groups are well organized in defining and prioritizing their needs, particularly resource-intensive fields that require coordination and consensus for instrument design, data collection and sampling strategies. Others, such as researchers in the life sciences, are more decentralized in their activities, in part because until recently, leading-edge science in these fields could only be pursued at the single lab level. And yet other groups are only beginning to understand, make use of and exploit DRI capabilities. These researchers need considerable support and training to make the most of DRI to advance their research.

Institutions

Universities, research hospitals and colleges have an interest in ensuring their researchers are well-served by the DRI ecosystem. They do this by providing research computing resources and services at the institution, as well as access to national DRI when their needs surpass institutional capabilities and expertise.

Institutions also provide an essential foundation of DRI in the form of desktop computers, departmental computing clusters, data storage servers and connectivity to the Optical Regional Advanced Networks. Researchers can scale up from these locally supported, institutionally based digital infrastructure to the shared advanced computing resources offered by Compute Canada and CANARIE when a growing need warrants it. While some of this infrastructure is acquired through external sources of funding, institutions remain an essential source of support for the operating and maintenance of these assets which are critical to the DRI ecosystem.

Service providers

Service providers focus on the network backbone and shared computational and data storage resources.
CANARIE: Canada's Advanced Research and Innovation Network (CANARIE) is a not-for-profit corporation that supports Canadian-based research, discovery and innovation by providing research and education communities in Canada with a high-speed network to transmit and share data. It does so in concert with the Optical Regional Advanced Networks, a set of 12 regional networks that are connected to form the national service. CANARIE receives its funding through a five-year contribution agreement with the Government of Canada.

CANARIE has successfully accomplished its mission of providing Canada with one of the world's most highly performing networks. In a 2014 evaluation of its contribution to CANARIE[3], Industry Canada concluded that data-intensive researchers in particular are highly dependent on the network to transmit large volumes of data globally, in a reliable, efficient and secure manner across all fields of inquiry. It also suggested that the network both facilitates collaboration and data transfer in Canada and across the globe, and makes digital content readily accessible.

……..

[3] Industry Canada, *Evaluation of Industry Canada's Contribution to CANARIE, 2014*

Compute Canada: Compute Canada is an incorporated not-for-profit organization that coordinates Canada's national advanced research computing resources for the benefit of academic researchers requiring advanced research computing resources within Canada, irrespective of their location or institutional affiliation. It is currently managed by a four-member executive team and governed by an independent board of directors. Membership is available to any university or college that has at least one researcher using Compute Canada's advanced research computing resources. Currently, 33 universities are members of Compute Canada. The organization works closely with four regional partner organizations — ACENET in Atlantic Canada, Calcul Québec, Compute Ontario and WestGrid — to ensure the optimal management and timely coordination of access to resources and support services for users.

Compute Canada's governance model is unique: it combines a federation of institutional and regional consortia with a national organization that reports to an independent board of directors. The activities of the board are supported by two independent committees: the Advisory Council on Research and the International Advisory Committee. The four regional partners are observers on the board.

While institutions own the advanced research computing infrastructure assets and employ the staff responsible for the operations and use of the computing resources, Compute Canada manages and coordinates the operation and access to these resources. The CFI contributes funding for both the capital and operational costs of the pan-Canadian advanced research computing platform.

Allied organizations and associations

Allied organizations and associations are membership-driven groups that represent several academic institutions and other stakeholders. These include the Leadership Council for Digital Infrastructure, the Canadian University Council of Chief Information Officers, the Canadian Association of Research Libraries and Research Data Canada. The mandates and activities of these organizations overlap to some degree.

Leadership Council for Digital Infrastructure: This council is composed of a group of university leaders who represent stakeholders with a part to play in the creation of a world-leading, advanced digital infrastructure ecosystem for Canada. In recent years, this has included Research Data Canada, the Canadian Association of University Chief Information Officers, the National Research Council Canada's Knowledge Management program (formerly CISTI), the Canadian Association of Research Libraries, Compute Canada and CANARIE. The Leadership Council has focused on:

- Developing a vision for Canada's advanced DRI ecosystem;
- Providing a forum for all DRI stakeholders to exchange views and coordinate activities;
- Providing opportunities for the community to determine current and future priorities and collective success factors for moving forward;
- Conducting an environmental scan of the current state of Canada's DRI and those of other jurisdictions; and,
- Developing a policy framework and roadmap to guide collective and concerted actions dealing with DRI and to ensure that future investments are both strategic and maximized.

The Leadership Council is a forum for interested parties seeking to reach consensus on a vision for Canada's DRI, assess how the DRI ecosystem is functioning and agree on priorities.

Canadian University Council of Chief Information Officers: CUCCIO is a non-profit, member-funded association of Canada's higher education information technology (IT) leaders, working together to help Canadian universities excel through the innovative and effective use of IT. CUCCIO is comprised of the chief information officers from more than 50 universities across Canada. Its strategic priorities include:

- Advancing best practices, sharing information, exploring new ideas and celebrating accomplishments;
- Collaborating with colleagues to achieve shared objectives; and,
- Having a strong and collective voice for the strategic role and contribution of IT to their institution and to the broader higher education community.

Canadian Association of Research Libraries: CARL is the leadership organization for the Canadian research library community, representing more than 30 of Canada's large university research libraries. Its primary mandate is to enhance the capacity of research libraries to advance research and higher education by promoting effective and sustainable scholarly communication through broad access to scholarly information. CARL members provide invaluable support to Canada's research community by:

- Working to improve access to knowledge
- Improving services to students, faculty and researchers
- Promoting effective and sustainable scholarly communication
- Ensuring that Canada's researchers have world class information management capacities
- Sharing best practices and experiences
- Advocating for public policies that enable broad access to scholarly information

Research Data Canada: Created in 2012, Research Data Canada is a stakeholder-driven organization dedicated to improving the management of research data in Canada. It brings stakeholders together to develop strategies, facilitate communication and partnerships, promote education and training, measure progress, and identify gaps. It also acts as a Canadian point of contact for international initiatives.

Since its creation, Research Data Canada has been supported by the National Research Council Canada's Knowledge Management program, CARL and CANARIE. To meet the need for greater coordination and promotion of research data management, the sponsoring organizations are seeking ways to ensure that Research Data Canada can play a larger role in the DRI ecosystem and enhance its activities.

Funders

The three federal funding agencies (CIHR, NSERC and SSHRC) traditionally have played a significant role in the structuring and funding of DRI and advanced training in Canada. The division of responsibility has evolved since the creation of the CFI in 1997 and of Genome Canada in 2000. Over the last 15 years, the CFI has gradually become one of the primary funders for the kind of advanced research computing infrastructure not typically provided by an institution to its researchers and those secured through research grants. The CFI has also encouraged consolidation of advanced research computing resources and services to more effectively optimize the system.

The granting agencies have focused on their core missions of providing research support to researchers, either individually or in small groups. They also collaborate in matters related to policy development, the most recent example being the joint draft policy statement on research data management.

Collaboration also extends to the joint support of research that uses DRI, with the most recent examples being: the 2014 partnership of the tri-agencies and CFI in NSERC's Discovery Frontier initiative called Advancing Big Data Science in Genomics Research; and, the 2013 Digging into Data initiative coordinated through SSHRC, which involved both NSERC and the CFI. Currently, CIHR and Genome

Canada are also collaborating on the development of a national strategy for bioinformatics and computational biology.

In 2006, the CFI launched the National Platforms Fund for high performance computing with the goal to enhance the country's DRI. It remained the only targeted investment in advanced computing research infrastructure until the CFI's new Cyberinfrastructure Initiative was launched in 2014. Over the past five years, an absence of predictable and periodic reinvestments in advanced research computing infrastructure led to a piecemeal strategy for enhancing the pan-Canadian advanced research computing platform. This meant that smaller-scale computing resources secured via other CFI-funded projects were added to the platform. While the opportunity to integrate and manage these smaller-scale computing infrastructure assets was welcome, the efforts required were significant, and in many cases burdensome.

The CFI launched its Cyberinfrastructure Initiative in 2014 to address two opportunities:

- To unlock the full potential of research data by organizing, structuring and integrating data sets into shared research data infrastructures, including developing analytical tools to mine the data; and,
- To upgrade and modernize Canada's advanced research computing platform in a way that supports computationally challenging research and data intensive research.

Provinces and regional economic development agencies of the federal government

Provinces play a critical role as a key funder of DRI in Canada. In some cases, this support is provided through provincial funding agencies, while in others it comes from departments of the provincial governments. Provincial support comes from the funding for postsecondary institutions for new computing infrastructure and its operating costs, along with their Optical Regional Advanced Networks. Provinces also provide partner funding to support successful CFI-funded research infrastructure projects. Some provinces — Alberta, Ontario and Québec, in particular — are developing their own DRI strategies. Some are also supporting initiatives that enhance the national DRI capacity, such as the Southern Ontario Smart Computing Innovation Platform — a collaboration between the province, IBM and 14 postsecondary institutions in Ontario that worked together to acquire Canada's fastest supercomputer. Significant funding for this initiative was provided through FedDev Ontario, one of the federal government's regional economic development agencies. Other such collaborations exist in British Columbia and Nova Scotia, and are planned in Quebec, each with significant funding from their respective federal economic development agency.

Although these investments provide substantial advanced research computing resources and support services, they currently operate as closed collaborations where access is granted only to participating institutions. These resources could be better integrated with the rest of the country's DRI if they were managed and operated as truly shared resources for institutions, communities of researchers and users from across the country.

APPENDIX 2: THE SEVEN ATTRIBUTES OF A HIGH-PERFORMING DIGITAL RESEARCH INFRASTRUCTURE ECOSYSTEM

- **Integrated:** the strategy ensures that all actors are integrated in a coordinated national system that minimises duplication, fragmentation and overlap, and avoids competition between stakeholders and service providers;

- **Inclusive:** the strategy draws information, insights and knowledge from all stakeholders;

- **Sustainable:** the strategy is supported in a way that enables evolution and adaptability, agility and responsiveness to scientific and technological changes;

- **Comprehensive:** the strategy provides the full spectrum of digital services and capabilities required by the Canadian research community;

- **Accessible:** the strategy allows for ease of access regardless of location, discipline, level of expertise or type of platform;

- **User-centric:** the strategy supports a system that is responsive to user needs and is focused on rapid and efficient service delivery; and,

- **Adaptable:** the strategy allows for emerging and evolving needs to be addressed in a nimble and timely fashion.

Research builds communities
La recherche au service des collectivités

450-230 Queen St.
Ottawa ON K1P 5E4
Tel 613.947.6496
Fax 613.943.0227

450-230 rue Queen
Ottawa ON K1P 5E4
Tél 613.947.6496
Téléc 613.943.0227