Workshop on cyber-infrastructure

January 23, 2014

Toronto, Ontario

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

WORKSHOP REPORT

On January 23, 2014, the Canada Foundation for Innovation hosted a workshop with 15 highly accomplished Canadian researchers on how best to support the development of cyber-infrastructure for research in Canada. The focus of discussion was on the major challenges facing researchers over the next several years and how cyber-infrastructure, including computational capacity, data management, high-speed networks and analytical tools, can be employed to help meet these challenges. The purpose of the workshop was to provide the CFI's senior management team with an opportunity to learn about the cyber-infrastructure needs of the Canadian research community, across a range of disciplines, research areas and institutions.

Context

At the CFI's request, the Government of Canada announced in the 2013 Federal Budget that a portion of the interest income accrued by the CFI will be allocated to support cyber-infrastructure at Canada's universities, colleges, research hospitals and non-profit research organizations. In accordance with the overall CFI mandate, these funds will be used to support cyber-infrastructure initiatives that assist eligible institutions to:

- Attract and retain the world's top research talent;
- Enable researchers to undertake world-class research and technology development that will lead to social, economic and environmental benefits for Canada;
- Support private-sector innovation and commercialization; and
- Train the next generation of researchers.

The CFI routinely funds eligible institutions to establish research infrastructure projects that incorporate various elements of cyber-infrastructure. Some of these projects develop common or shared cyber-infrastructure resources, but that is generally not the primary intention of the researchers involved. For this initiative, the CFI is considering an alternative approach, in which the cyber-infrastructure supported would primarily be intended as a common or shared resource, available for use by a broad range of researchers across institutions and disciplines.

Seeking input on how best to design and deliver a cyber-infrastructure initiative, the CFI asked the workshop participants to consider and reflect upon the following questions:

1. What scientific and societal challenges will researchers face in three to five years that can best be addressed through the development of common or shared cyber-infrastructure?
2. In light of the opportunities on the horizon, what types of cyber-infrastructure investments (hardware, software, personnel, services and organizations) are likely to make a meaningful, lasting and structuring impact on the research community's ability to address these challenges?
3. Given the scope of the CFI's mandate, are there specific priorities that should be taken into account in making these investments?

What we heard:

‘The nature of research is changing’

Massive amounts of data, enormous computational capacity and light-speed communications networks that link Canadian researchers to the world are fundamentally transforming the way research is conducted, the nature of the questions being asked and the impact that new knowledge is having in all areas of society. The emergence of what has been called the fourth paradigm of research, where the collection, organization, visualization and analysis of digital information drive the formation of hypotheses, is increasingly becoming the norm across numerous areas of research in Canada.

This transformation is taking place in research areas as diverse as the humanities and high-energy physics, environmental modelling and clinical medicine. This dynamic presents the main challenge in supporting cyber-infrastructure, ensuring that the research community as a whole has the computers, software, storage devices, networks and data management expertise necessary to take full advantage of computationally intensive, digitally enabled modes of inquiry.

‘We’re reaching the limits of current capacity’

Although Canada’s current cyber-infrastructure environment is, in most instances, sufficient to support the work of academic researchers across the country, bottlenecks are emerging in some areas, and particular elements, such as training and data archiving, have not yet been adequately addressed. For example, in certain areas of physics, such as turbulence modelling in ocean and atmospheric systems, a single calculation on some of the fastest computers in the country can take several months to complete. In other areas, the science is limited not by the ability of instruments to generate data but by the current capacity of computational, storage and data processing platforms. At the Atlas Canada Tier-1 Data Centre, the volume of data could be increased by a factor of five, but the computational and storage capacity of the platform simply cannot handle this degree of expansion.

The use of digital content in practically all areas of research is creating challenges for the various elements of Canada’s cyber-infrastructure environment. While generally adequate to meet today’s needs in most areas, the national high-performance computing network, managed by Compute Canada, is rapidly depreciating and is not necessarily designed to meet tomorrow’s research challenges. As one participant noted, there is a growing need for different types of high-end computing equipment; it is not just a matter of replacing what we have with more of the same. Moving data and, in particular, very large datasets is also rapidly reaching the limits of the current infrastructure. And although CANARIE and the provincial ORANs are capable of expanding network capacity to meet researchers’ needs, there are real limitations at the institutional level (i.e. the last-mile issue). A key aspect of addressing this challenge is the need to develop a broad range of middleware and software products capable of organizing, categorizing and analyzing today’s large-scale and often highly complex datasets.

‘Data are getting larger and more complex’

The drive to create new knowledge from data is increasing both the volume of data collected and its complexity. Researchers in numerous areas are now seeking to combine digital information from a wide variety of sources to conduct completely new types of analyses. For example, using the CBRAIN platform, researchers are combining genetic data with psychological profiles and neurological imaging to conduct multi-variance analyses of individual patients’ conditions. Others are exploring phenotypes

through the combination of genetic data and clinical health data to develop personalized-care strategies. These, and many other examples, point to the need for integrated computational, storage and software systems that have the capacity to handle heterogeneous data that come in various scales and formats.

‘Continuing need for experimentation and technology development’

Many of the digital technologies currently in use by Canadian researchers are in relatively early stages of evolution. Computational capacity continues to expand rapidly. Data storage is generally keeping pace, while costs have decreased dramatically. Both generic and highly specialized analytical platforms are now being developed in areas such as systems biology, neurology and the humanities. There’s a long way to go, however, in matching technological ability to the fundamentally important questions that today’s researchers are exploring. This creates both barriers to scientific progress and enormous potential in technology development. As a number of workshop participants suggested, the best way forward is to foster closer connections and greater integration between the people seeking to solve scientific challenges and the developers of the tools needed to do so. There is, however, no need to reinvent the wheel; Canadian researchers and their institutions can learn a great deal from advances taking place outside the country. And although in some areas, such as imaging and high-energy physics, Canadians are global leaders in systems development, we are falling behind in others.

‘The unique Canadian context and characteristics can be coupled with international experience and expertise’

The Canadian research community benefits from a unique combination of social, legal and technological characteristics that can be effectively built upon to expand and improve the overall cyber-infrastructure environment. As one participant noted, the Canadian research community is large enough to have an impact at the global level and yet small enough that it can be efficiently organized. As well, our legal and ethical frameworks in areas such as data confidentiality are widely respected around the world. Canadian researchers also have privileged access to unique geography and a multilingual, multicultural population. However, we can’t do everything, and the key challenge will be to identify both existing and emerging strengths that can be built on to be globally competitive. In many areas of research, doing so will require drawing on and learning from international experience and expertise.

‘Potential of generic platforms to support specialized capacity’

In the development of cyber-infrastructure platforms, there is an inherent tension between the generic and the specialized. Generic platforms, such as Compute Canada, can and do meet many computational needs across a wide spectrum of the research community but, by the nature of their design, cannot meet all needs, especially in areas where highly specialized functions are required or where there are legal and ethical constraints around the use and sharing of particular types of data. This tension highlights a number of particular issues that require consideration in the allocation of resources necessary to build a leading-edge cyber-infrastructure environment. At what point do economies of scale in areas such as data storage begin to break down? What are the limitations of middleware in linking and facilitating the analysis of disparate datasets? On what basis do we determine which data should be archived for long-term preservation and which should be discarded? It is increasingly evident that these questions can be at least partially addressed through an appropriate combination of both generic platforms and specialized capacities. Ensuring ongoing coordination between researchers, technology developers and platform managers will be necessary to ensure success.

‘Training and the development of expertise across all areas are crucial’

The workshop participants repeatedly stressed the need for training and expertise development across all areas of cyber-infrastructure: data management and curation, analytics, computer engineering, middleware development and the provision of services. As participants stressed, the difficulties in training and retaining high-level expertise — people power — are, to some extent, compromising the effectiveness and efficiency of cyber-infrastructure capacities; people leave, data are lost, research teams are broken up, and researchers are left unable to access systems, change computer codes and explore new capabilities.

The need to develop expertise across a wide range of disciplines and in sufficient numbers to meet demand is, in itself, a rationale for a distributed cyber-infrastructure environment. Effective training requires local on-the-ground contact with researchers, support from institutions and challenging intellectual environments. The required numbers of highly trained experts alone suggest that a centralized approach will be insufficient. Participants indicated that this is one area where coordinated action between the granting agencies and the CFI would be beneficial and perhaps should be considered a priority.

‘Potential for working with industry, but cultural barriers exist’

In today’s research environment, researchers don’t just collect data themselves; increasingly, they draw on a wide variety of types of digital content from other activities, organizations, institutions and areas of human endeavour. Economists, sociologists and public health researchers regularly use large-scale datasets from statistical agencies, regional health authorities and private-sector companies. This presents a particular set of challenges, many of which have less to do with technical capacity than with cultural, legal and operational realities. For example, the Statistics Canada data made accessible through the university-based Research Data Centres can be used only by researchers who agree to swear an oath of confidentiality and can be accessed only within the centres themselves. Comparable census data from the United States are all but impossible for Canadian researchers to access. Researchers working with the private sector often find that significant cultural differences between academic research and companies create real barriers to research, even when the results could potentially benefit the companies directly. These types of barriers are difficult to overcome, even more so than technical barriers, but credible and respected regulatory regimes, along with effective data security systems, can be useful.

‘There are real challenges in coordinating funding, expertise and systems development’

It is perhaps not surprising that the combination of the ubiquitous use of cyber-infrastructure across practically all disciplines, the broadly distributed nature of the Canadian research community in universities and colleges, both large and small, and the numerous granting agencies at two jurisdictional levels has produced a situation where coordination is a major concern. The workshop participants noted the need for greater coordination at all levels: between general service platforms like CANARIE, Compute Canada and specialized disciplinary or institutional initiatives; between computer scientists who develop cutting-edge technologies and researchers who want to put these technologies to use; between institutions that have researchers working in the same areas who need common tools to collaborate effectively; and between granting agencies at the provincial and federal levels that provide the needed financial support. The participants, however, did not suggest that some sort of new organization or institution is necessary to ensure coordination. Rather, the development of both general and specialized platforms built around broad areas of research would, in itself, foster coordination and collaboration.

‘Compute Canada has a limited capacity to integrate specialized project-level equipment’

While the high-performance computing capabilities managed by Compute Canada are considered absolutely vital to the ability of Canada’s researchers to create new knowledge, there are limitations to what Compute Canada can and should do. Compute Canada is effective at reducing the duplication of very expensive computational resources and is becoming proficient at computational resource allocation, but it does not have the capacity to develop or house other elements of the cyber-infrastructure system. Middleware, for example, should be developed in coordination with Compute Canada so that it runs effectively on Compute Canada’s computing systems, but ultimately, the middleware should be attached to the data and the researchers who use the data. This way, it can be designed to better meet the needs of researchers rather than those of the machine operators. In certain instances, however, it may be productive to have Compute Canada bundle similar individual project-level systems and make them available to a broader range of researchers. It all depends on the degree of generalizability.

‘Tension exists between excellence and building common resources. Should elements of cyber be funded as a general utility?’

One of the key questions that workshop participants raised is the inherent tension between supporting individual initiatives through an excellence-based competitive process and supporting those elements of the cyber-infrastructure system that are, by their very nature, general-use utilities. Compute Canada, for example, is a single unique structure that does not have an equivalent comparator and, therefore, does not lend itself to a competitive process. On the other hand, individual platforms that operate on top of the Compute Canada structure could be supported through a competitive process. The participants advised, however, that these platforms should be linked, wherever possible, with the general-purpose facilities and operate at a higher level than an individual laboratory. A useful example in this regard is the Atlas Canada Tier-1 Data Centre and its functional partnership with SciNet, one of the Compute Canada nodes. As was pointed out, this sort of partnered relationship fosters specialization, meets the cyber-infrastructure needs of an area of research (in this case, high-energy physics experiments conducted out of CERN) and diminishes unnecessary duplication.

The key for the CFI is to ensure that research excellence is the driver and that this is determined through a competitive process. The challenge is to ensure that linkages are established between the general-utility infrastructure services provided by Compute Canada and CANARIE and the specialized functions that can serve researchers across multiple disciplines.

‘Periodic, uneven funding limits consistent capacity building’

On several occasions, the workshop participants pointed out the limitations of current funding models and the negative impact such models have on the research community’s ability to consistently build cyber-infrastructure capacity. The CFI’s periodic and unpredictable funding model creates difficulties in retaining highly qualified expertise, in refreshing platforms with new technologies and required capacities and in addressing particular areas such as data archiving, where long-term stability is a necessity. The division of funding responsibilities between the federal granting agencies, the CFI and specialized agencies such as Genome Canada means that it is often difficult to line up and coordinate funding streams. At times, the direct research support is in place while the infrastructure funding is not. At other times, it is the reverse. In some areas, such as health genomics, funds are available but cannot be used to support general platforms. In other areas, such as the humanities, funding for the expertise needed to support cyber-infrastructure developed to meet specific researcher needs is lacking. The participants pointed out that in

some instances, interagency funding initiatives, such as Discovery Frontiers and Digging into Data, have proven useful but should not be considered a one-size-fits-all solution. There are simply too many differences in the needs and levels of development of various research areas for this approach to be universally effective.

‘Build a common backbone and let more specialized initiatives operate on top of it’

The participants did not attempt to reach a consensus on particular issues or possible ways forward, but in general, they advised the CFI to fund both the common backbone, such as Compute Canada, and specialized initiatives that develop tailored cyber-infrastructure capacities for distinct areas of research. These specialized initiatives would then be able to provide particular functions, such as customized software development, large-scale data management, data archiving, security and confidentiality protections or whatever capacity or combination of capacities is needed in a specific area of research. Such an approach would also accommodate differences in levels of cyber-infrastructure development that exist between disciplines and areas of research. As one participant put it: Although the main highway is now in place, some still need an on-ramp, one that is best designed by the researchers themselves. Others who are already on the highway need the means to go farther, faster.

‘Let competition criteria drive selection of fundable initiatives’

In the end, the participants suggested a number of specific measures that the CFI should consider, including continuing to support the backbone services, seeking out funding partnerships with the federal granting agencies, where appropriate, and distributing currently available funds through multiple competitions to smooth out investments. Funding should be allocated to each stream according to clearly identified needs and to the future potential of supporting research excellence.

The competitive elements, however, should not be overengineered. The best approach is to offer open competitions with clear and well-established criteria: excellence, applicability, the training of highly qualified persons, potential for innovation and demonstrated need. Given that the CFI routinely funds project-level cyber-infrastructure, criteria of common or shared infrastructure to meet the needs of areas of research should also be incorporated into the adjudication process.

Research builds communities
La recherche au service des collectivités

450-230 Queen St.
Ottawa ON K1P 5E4
Tel 613.947.7260
Fax 613.943.0227

450-230 rue Queen
Ottawa ON K1P 5E4
Tél 613.947.7260
Téléc 613.943.0227