June 2014

ADVANCED COMPUTING TRANSFORMING INNOVATION IN ONTARIO (ACTION)

Needs and Opportunities for Advanced Computing in Ontario

orion

TABLE OF CONTENTS

orion

Ontario’s Backbone of Innovation.

ADVANCED COMPUTING TRANSFORMING INNOVATION IN ONTARIO (ACTION):

A Report on the Needs and Opportunities for Advanced Computing in Ontario

FOREWORD

ORION initiated the Advanced Computing Transforming Innovation in Ontario (ACTION) project to strengthen our understanding of the needs of Ontario's research, education and innovation (RE&I) communities as they relate to advanced computing (AC).

Why would ORION—as the backbone of innovation for Ontario providing RE&I broadband connectivity—be interested in AC?

The answer is quite simple: we connect universities, research hospitals and research institutions—and all of them have used or will use AC. With more than half a million researchers using our network to complete their work, it is important for ORION to understand their AC needs in order to ensure that the required connectivity and the next generation of network-enabled services are available to them.

The ACTION project focuses on the needs of researchers as the end user. It also examines the relationship of AC to other elements of digital infrastructure, beyond computing and connectivity. This includes data, cloud services and collaboration. In today's world, these elements are best addressed as a complete ecosystem and not as individual items. To help maintain this focus, this report examines these issues through the lens of researchers' AC needs.

That information was gathered through in-depth, face-to-face meetings with a broad cross-section of Ontario's researchers. It was further supplemented with information gathered at a community workshop held in Toronto in February 2014, along with a broader consultation with Ontario's research community. It is that collective, community input that feeds the outcomes of the ACTION project to date.

As part of the ACTION project, this report takes the reader through the analysis process to make specific recommendations for short- to medium-term collaborative activities and programs that can and should be undertaken to address the needs and opportunities of Ontario's RE&I

communities. This has been an extremely worthwhile initiative in its own right, but it doesn't end here. The next task will be to see how some of these ideas can be implemented to benefit Ontario's researchers.

ORION greatly appreciates the work of Bill Appelbe, ACTION's Project Director and the lead author of this report. We would also like to thank the researchers who participated in this study, Ontario's Ministry of Research and Innovation and the members of the technical working group,[1] who represent a broad cross-section of expertise in the research and advanced computing sectors in Ontario.

Darin Graham, PhD
President & CEO, ORION

[1] Further information about the technical working group can be found in Appendix A.

Executive Summary

Research and innovation are increasingly reliant upon advanced computing (AC). This is widely acknowledged and has been documented in numerous reports and studies conducted at provincial, national and international levels.[2]

Despite this increasing dependence, Canadian investment in AC has fluctuated widely over the years. Most reports and studies have concluded that Canada is falling behind in its investment in AC infrastructure, particularly with respect to high-performance computing (HPC). HPC involves large computing systems with thousands of processors capable of modelling complex scientific applications or analyzing immense data collections.

What has not been adequately addressed are the requirements and support researchers need that extend beyond AC infrastructure. For example, although reports have identified that there is a strong demand for highly qualified personnel (HQP), the scope of this demand has not represented a serious area of inquiry. However, HQP are needed to support researchers in their use of AC, and limited access to HQP increasingly limits researchers' ability to effectively use AC infrastructure.

Other trends in research and information technology (IT) are also having a significant impact on the AC landscape, including the growth of big data and cloud computing, as well as the increasing interdependence between healthcare and bioinformatics research and development (R&D). Nonetheless, previous studies and investment proposals have generally ignored these trends.

By contrast, this report seeks to address these needs and opportunities by conducting a detailed analysis of the productivity and competitiveness of a broad sample of researchers in Ontario.

The conclusions drawn from this analysis can be summarized as follows:

1. Researchers' dependence on AC is rapidly increasing, to the point where nearly all modern research requires AC.

2. Traditional HPC is only used by, and effective for, a small fraction of researchers. Many researchers need more interactive, adaptable and on-demand computing than is provided by traditional batch HPC.

[2] Please see Appendix D – References for a list of studies reviewed as background for and referenced in this report.

3. Access to and management of research data is increasingly important to researchers, but is not adequately addressed by the existing AC support services and facilities.

4. There is a shortage of accessible HQP supporting researchers' effective use of AC. More than half of the researchers interviewed indicated a need for greater HQP support.

5. In Ontario, there is a complex and often confusing checkerboard of entities and organizations providing AC support and services, including networking, storage, computing and applications support. Few manage to navigate this confusing landscape successfully.

6. The barriers to HQP and AC access are limiting the productivity and competitiveness of research conducted in Ontario and Canada.

To address these conclusions, this report makes the following recommendations:

- Since Ontario's investment in AC infrastructure and HQP development is in arrears, more investment is required, including that which would provide a broader base of AC services than traditional HPC. This could come about through means such as provincial shared services, cloud computing, a combination of local and cloud computing, and data management.

- A broad-based provincial strategy should be implemented in order to develop HQP and connect researchers to existing HQP expertise.

- Information about AC support and services in the province should be made more accessible. Ideally, this information could be made available through an information website.

A Definition of Advanced Computing

In analyzing researchers' dependence upon advanced computing (AC), it is important to start with a clear understanding of what AC encompasses. While many definitions of AC exist,[3] this report uses the term to refer to any specialist IT software or hardware that is not widely available to the general public (as opposed to common word processing or Internet tools) and typically requires expertise or highly qualified personnel (HQP) to utilize and support it.

Advanced computing includes traditional high-performance computing (HPC), which is supported by large-scale computing clusters. However, most current AC use by researchers is not HPC. Instead, it may rely on large-scale storage, visualization, departmental clusters or specialist desktop computing systems.

It would be an oversimplification to view AC merely as the IT infrastructure (for example, computers, storage, networks and software) that researchers use to conduct their research. Following this line of reasoning, researchers only need sufficient access to AC infrastructure—just as they need sufficient access to commodity IT tools, email support, word processing and Internet access.

The reality is very different: there is a symbiotic, collaborative relationship between AC and research. It is a community or ecosystem. According to this study and previous reports, better access to AC facilities enables better research—tackling larger and more complex problems, analyzing larger data sets, collaborating more effectively, resulting international recognition and more research funding. Advanced computing is a key enabler of modern research, not just a tool. It means far more than the mere access to infrastructure—AC is the combination of HQP and infrastructure that researchers need to be productive and competitive.

Support from HQP for AC covers a broad spectrum of specialist IT skills, including systems administration, software applications support, programming and technical analysis. There is a strong demand from researchers for support staff with these skills. However, acquiring these specialist IT skills typically requires both an undergraduate degree and on-the-job training and mentoring. Providing HQP support for researchers requires a broad set of skills, and AC support groups are most effective when they include a collaborative team of specialists. Connectivity and collaboration among AC support staff exists informally, but strengthening it provincially could be extremely beneficial, especially for the majority of institutions and researchers that do

[3] Compute Canada defines AC as "the portfolio of expertise, services, customized software solutions and high performance computational capacity that enables excellence in research, research training and innovation." (https://computecanada.ca/cc_files/news/CC_NationalConsultation_StrategicPlanDraftOutline.pdf)

not have access to a local support team of HQP. In almost all cases, modern AC platforms and tools require researchers to have access to HQP in order to adopt and use them effectively.

Advanced Computing in Ontario

Ontario has three well-established and highly regarded HPC centres: the High Performance Computing Virtual Laboratory (HPCVL), SHARCNET and SciNet. Together, these three centres have approximately 60 HQP support staff, who are distributed across a dozen universities. There are also strong HPC support groups affiliated with other organizations, particularly among the health research institutes—for example, the Ontario Institute for Cancer Research (OICR) and HPC4Health, the latter of which involves collaboration between the Hospital for Sick Children and the University Health Network (UHN). However, provincial collaboration among these organizations, as well as others providing AC services and support in Ontario,[4] should be strengthened to provide researchers with easier access to AC support and services. Compute Ontario's formation could be the catalyst that improves this situation.

In addition, Ontario is home to about 47% of all Canadian research, according to 2014 statistics from the Ontario government. This research is conducted within dozens of independent universities and research institutes that are spread over a wide geographical area. Expertise and information is thus highly distributed—a characteristic that exists in few other jurisdictions within Canada.

The Canadian government's priorities for science and technology confirm that leading-edge research is dependent on AC. Canada's 2007 Science and Technology Strategy[5] identified four research areas where Canada can leverage its strengths to achieve a competitive advantage:

1. *Natural resources and energy.*

   Sub-priorities: Energy production in the oil sands; Arctic resource production, climate change adaptation and monitoring; and biofuels, fuel cells and nuclear energy.

2. *Health and related life sciences and technologies.*

   Sub-priorities: Regenerative medicine; neuroscience; health in an aging population; biomedical engineering and medical technologies.

[4] A partial list of AC and research organizations is included in Appendix A. Some examples include the Ontario Centres of Excellence (OCE), the Southern Ontario Smart Computing Innovation Platform (SOSCIP) and ORION.
[5] http://archive.nrc-cnrc.gc.ca/eng/news/nrc/2008/09/02/st-strategy.html

3. *Information and communications technologies.*

    Sub-priorities: New media, animation and games; wireless networks and services; broadband networks; and telecom equipment.

4. *Environmental science and technologies.*

    Sub-priorities: Water (health, energy, security); cleaner methods of extracting, processing and using hydrocarbon fuels; and the reduced consumption of these fuels.

Without exception, these research areas depend upon AC for data analysis, optimization, forecasting, modelling and simulation. These federal priorities in research areas align with many of Ontario's priority areas, such as healthcare and the life sciences.

It is also important to note the socio-economic benefit of these research areas. In particular, there is a growing interdependence between bioinformatics research and clinical practice. As such, having a strong strategy to improve access to AC could significantly benefit clinical practice, health outcomes and healthcare costs. Conversely, inadequate AC infrastructure that fails to support R&D strategies will have negative effects and impose higher costs—not only on the research sector, but also on the health sector, which comprises a large component of the provincial budget and GDP.

Countless examples illustrating the importance of AC and HQP to research in Ontario and Canada have been cited in many previous studies (see Appendix D). As a result, AC infrastructure and HQP are now being recognized as important assets and resources in themselves, not just as a cost or overhead for conducting research.

Examples include OICR, which is recognized for its data collections and software, in addition to its research. Similarly, the Ontario Brain Institute (OBI) is recognized for its Brain-CODE software platform, a data repository of various brain disorders that allows researchers to explore the complex relationships between disorders in order to identify new methods of treatment.

However, simply identifying the importance of AC does not answer questions regarding issues such as the depth and breadth of that dependence for specific research disciplines, whether current support is adequate, and why that is or is not the case.

Some of the needs identified in this report amplify needs that have already been noted and investigated (for example, in the 2009 report *Cyberinfrastructure and the Research Process in Canada* and the 2012 report *Research Computing in Ontario – A Framework for the Next Generation*). However, research and AC continue to evolve rapidly, and the amount and diversity of research dependence on AC is greater today than ever before, in part due to emerging trends such as big data analytics, cloud computing and the explosive growth in biomedical data.

This report seeks to add an in-depth analysis of current researchers' needs to the existing body of knowledge, and to provide recommendations for the short- to medium-term measures that can be taken to address these needs. This includes a quantitative analysis of the extent of AC needs, extracted from in-depth meetings with a broad spectrum of researchers, and specific recommendations on how these needs might be met.

Survey Methodology and Goals Summary

In analyzing and understanding the use of AC in Ontario, this report focuses on five fundamental questions:

1. What proportion of researchers use AC resources and to what extent?
2. What types of resources are being used: computing, data or HQP?
3. Is the use of and dependence on AC forecast to increase?
4. How adequate is current AC support?
5. What steps can be taken to improve access to AC resources?

To answer these questions, face-to-face interviews were held with about 50 leading research groups across seven Ontario universities and research institutions. These interviews typically lasted 30 to 60 minutes.

The researchers interviewed work in a broad spectrum of fields, including mathematics, economics and the humanities, as well as in areas that traditionally use HPC, such as science and engineering.[6]

While the quantitative data from these meetings forms the backbone of this report, it should be noted that observations from informal meetings with additional stakeholders and relevant reports have also been included. Further information on the methodology used to interview researchers, as well as a copy of the questionnaire form, can be found in Appendix B.

There are also several important caveats that must be declared concerning the data gathered in this report.

First, this report and analysis is not intended to be an institutional review. For example, researchers were not asked to rate or rank their institutional support for AC. However, researchers invariably commented on their experiences with institutional AC support at various levels, whether departmental, institutional, provincial or national. That information has informed some of the conclusions concerning the best practices for AC support proposed in this report—even though this report avoids any type of ranking or scoring of institutional AC support.

Secondly, the size of the sample and the interview process preclude any detailed statistical conclusions. Instead, clear trends were drawn from the information and experiences related during the interviews. For example, an

[6] See Table 1, below.

overwhelming majority of the researchers expect their use of AC to increase in the next few years. However, determining whether that overwhelming majority might be 99% or 95% would require a much larger, randomly chosen sample. Additionally, quantifying the increase of expected AC use is difficult because of the exploratory and unpredictable nature of research and the intrinsic uniqueness of each research project.

The data and statistics below are based on the number of research groups interviewed (not individual researchers), and the results were not weighted by the number of researchers. It should also be noted that all percentages reported here were rounded to the nearest 5%, due to the relatively small sample size. Furthermore, the feedback from researchers was often qualitative rather than quantitative, for example, in describing their needs for HQP. Qualitative data is often easier to obtain than quantitative data (such as ranking the need for HQP on a scale). Thus, the quantitative data reported below was generally indicated only on a binary scale of need versus no need, or used versus unused.

A larger survey would have been desirable, both to get more statistically significant data and to meet a broader spectrum of researchers and institutions. Unfortunately, time and logistics precluded this option. Moreover, alternative methods, such as email or online surveys, would likely have resulted in a very limited response rate and low quality of information. As a result, in-depth interviews were chosen as our primary method of gathering information, as they can often reveal greater insight into the underlying needs of a researcher.

Finally, this report is not intended to decide on or rank technology solutions or infrastructure investment, nor does it propose funding mechanisms. Instead, this report focuses on specific recommendations for collaborative initiatives that can be undertaken without major new investments.

Interview Results and Conclusions

1. Disciplines

Interviews with researchers did not find a single research group that was not dependent on AC in some way, highlighting its vital contribution to modern research. Of course, there are probably many researchers and faculty in universities and colleges in Ontario that are not presently using AC, and a smaller number that have no need for AC. Nonetheless, the evidence is overwhelming that a majority of active researchers—especially those who are regarded by their institutions as leading researchers—are highly dependent upon AC.

This survey was deliberately constructed to include a broad spectrum of research disciplines rather than being limited to only those that traditionally use HPC. Among the research groups interviewed, the distribution of disciplines was as follows:

Research discipline	*Number of research groups interviewed*
Economics	2
Humanities/Arts	3
Mathematics/Statistics	5
Computer Science/ Information Technology	6
Engineering	7
Science	9
Biomedical	9

Table 1 – Research Group Distribution by Discipline

The classification and distribution of research groups interviewed is intrinsically imprecise. For example, engineering overlaps computer science, and biomedical researchers frequently have faculty appointments in science departments.

2. Areas of Advanced Computing Need

At a high level, this study also examined the precise AC needs and opportunities in research communities. For example, do those needs include computation, data and/or HQP? How does that relate to the resources currently used by Ontario's research communities?

The results of these queries are as follows:

	Currently used and important	***A current "shortage" or a significant current constraint***	***Forecast to be more important in future (future need/ shortage)***
Data access and management, including security and privacy	75%	25%	70%
HQP (not including HQP within the researcher's current team)	60%	35%	55%
AC computing—local clusters or Compute Canada shared facilities	80%	30%	35%
Use of any AC resource (data, HQP or computing)	100%	80%	100%

Table 2 – Research Groups' Current Advanced Computing Use and Needs

Several observations can be made from the above data. Firstly, most researchers use more than one type of AC resource: computation, HQP and/or data storage and management. Data access, storage and management is as important as computational resources overall, although each individual researcher's needs are unique. For some traditional computational scientists, data management is not an issue. However, in disciplines such as bioinformatics, both computation and data management are important. In still other disciplines, such as economics or digital humanities, data access and management may be more important than large-scale computation.

The importance of HQP is indicated by the percentage of researchers who rely on AC support from outside their research group (60%). The source of that support is analyzed in greater detail below.

Finally, regarding future needs, it is interesting to note that data access and management are indicated as the highest priority (70%), while the growing demand for HQP is a secondary priority (55%). Lastly, the lowest ranking was given to computing cycles (or more flexible access to computing cycles) at 35%. It should be noted, however, that each individual research team and project is different. Some researchers are highly impacted by not having access to sufficient computing cycles, while for others the largest concern was gaining better access to HQP.

3. Support

The next major question investigated was the source of AC support that each research group relied upon, and whether it was provided at a local, institutional or national level:

Source	*Provision of HQP support*	*Provision of AC infrastructure*
Research group: the researcher and local team	70%	50%
Departmental and institutional hosting	35%	35%
Compute Canada's shared facilities	35%	35%

Table 3 – Sources of Researchers' AC and HQP Support and Facilities

Many research groups receive support from several sources. A key observation is that researchers receive about 50% of their AC infrastructure locally, typically from dedicated local clusters, which are then supplemented by institutional or national facilities. Most groups (70%) have their own local HQP to supplement the support that is available institutionally or from Compute Canada. However, a significant percentage (30%) do not have local HQP support available, and are thus reliant on institutional support or support from Compute Canada. As with most classifications, these are unfortunately a little imprecise (for example, the distinction between what qualifies as "local" or as "HQP"). Every institution and research organization has a different structure, rendering it difficult to be more precise. In addition, the researchers surveyed were generally considered to be leading or very active researchers at their institution, and as a result, were more likely to be better funded and have their own resources for AC.

While further analysis of the breakdown of researchers' needs with regards to computation, data and HQP can be found in Appendix C, this study draws the following high-level conclusions:

Researchers' dependence on AC is increasing.

Research dependence on AC runs deep. A large majority of researchers interviewed use AC on a daily or regular basis, and most of their research relies on AC services. As shown in the forecast column of Table 2, many of the research groups interviewed expect that their research dependence on AC resources (whether data access, HQP or computation) will increase over time.

Lack of effective access to AC is widespread and hinders researchers' productivity and ability to compete.

When asked about their access to AC (including infrastructure, such as hardware and software, as well as expertise and HQP), only 20% of research teams said that it was currently sufficient to meet their research needs. AC needs varied greatly across the different research teams, from requiring access to more HPC cycles to data management and, especially, access to HQP.

Support from HQP is highly valuable and in short supply.

Researchers highly valued AC support staff when they were available, especially local support. Local AC support staff can greatly increase the productivity of researchers in many ways, including assisting researchers in accessing and using the most appropriate AC tools and technology, providing support for local AC infrastructure and software, and even helping them write grant proposals. Lacking such local HQP support, researchers and their students often have to fend for themselves, which is both inefficient and unproductive.

It is obviously impractical to put local AC support staff at every research institution or department in Ontario. Internet tools and resources are a generic and increasingly popular alternative to face-to-face support and services. The question of whether online collaboration tools and support can lessen the need for local AC support, or make that local support more effective when it is available, is addressed later in this report.

Researchers with better AC access are more productive, better recognized and better able to attract funding.

Eighty percent of the research groups interviewed said they felt constrained to the point of having to tailor their research to the AC resources available to them (whether that was infrastructure, support or HQP), and/or that AC access had a significant impact on their productivity, ability to be internationally competitive, and even in their ability to attract funding and students to assist in their research. It can be argued that researchers always want more support, whether in funding, more or better qualified graduate students or teaching relief. However, discussions with researchers drilled down and documented the precise impact that the lack of AC access has on their research, ascertaining that it is a serious impediment to many researchers in Ontario. This is significant, despite the small sample size of this survey.

It would be extremely challenging to quantify the impact the lack of effective AC access has on researchers' productivity and competitiveness without conducting some rather invasive experiments (for example, providing similar researchers with differing access to AC for a few years to measure the consequences of that deprivation). However, the informal evidence gathered from this survey indicates that a lack of access to AC has a significant effect on a researcher's productivity and competitiveness.

The impact of that lack of access can be extrapolated in several ways from researcher interviews, for example, by quantifying the following:

1. The unproductive time researchers have spent mastering or learning to use AC and administering AC systems.

2. The lost opportunity costs resulting from the inability to tackle research problems due to a lack of AC resources, such as insufficient HQP access, computing, or data storage or management.

3. The inability to be competitive (for example, in attracting funding, postgraduate students or post-doctoral fellows). Several researchers commented that a lack of AC resources rendered it more difficult for them to attract research students compared to their colleagues in the US. Personal experience confirms that students and post-doctoral researchers are highly selective and more attracted to well-resourced research groups. Several researchers also indicated that having HQP assistance in writing the IT components of grant proposals helped make those proposals more

competitive; in fact, the heads of Canadian funding agencies have also commented on the importance of convincing and informed IT content in grant proposals.

In essence, the interviews confirmed that researchers with better AC access were more productive, better recognized, and better able to attract grants and funding. Conversely, researchers with poor AC access were constrained in their research.

Based on researcher interviews, it is estimated that the cost or value of better AC support is approximately one day per week—that is, 20% productivity loss/gain. If this estimate is accepted, then there are significant implications for institutional allocation of resources and competiveness.

It can be argued that it is a researcher's responsibility to locate AC support under a "survival of the fittest" strategy. However, some institutions do a significantly better job of supporting the AC needs of their researchers than others, which arguably increases the ability of those institutions to attract leading researchers and support their career growth.

The Impact of Limited Advanced Computing Access on Ontario's Researchers

About 35% of the research groups interviewed said they were constrained by access to HQP, while about 30% were constrained by access to AC or HPC infrastructure (see Table 2, above).

Need 1: Exploiting AC requires HQP, and a lack of access to HQP is even more limiting to researchers in Ontario than lack of access to HPC infrastructure.

Funding agencies at provincial and federal levels in Canada have traditionally tended to focus primarily on funding infrastructure and only secondarily on funding HQP. However, governments in Canada are now starting to recognize the increasing importance of funding HQP. Internationally, there are large-scale efforts to fund access to HQP and the development of platforms that make AC more accessible to researchers. For example, the XSEDE project, funded by the National Science Foundation (NSF) in the US, is a five-year project costing US$100 million, with more than 100 full-time employees.[7]

Advanced computing infrastructure varies from departmental clusters and storage to the Tier-1 national supercomputers with more than 100,000 processors (cores). In Ontario, the largest current HPC system is the Blue Gene Q system at SOSCIP, with about 40,000 cores.

In the AC ecosystem, it is important to recognize that researchers do not simply jump from using a desktop computer or application to using a supercomputer or specialist AC applications. The typical migration path first includes taking applications from a desktop computer onto a departmental cluster (up to about 10 cores or processors) when the research problem's computational or data needs exceed the capabilities of a desktop computer. This invariably requires HQP to support the migration: the research group either develops and gains the migration expertise itself, or receives that support from specialist AC staff. If the research group's needs for AC outgrow its current platform, a further migration is needed, and so on.

This can be illustrated as a pyramid, as seen in the following diagram.

[7] http://www.orion.on.ca/wp-content/uploads/2014/03/John-Towns-presentation.pdf

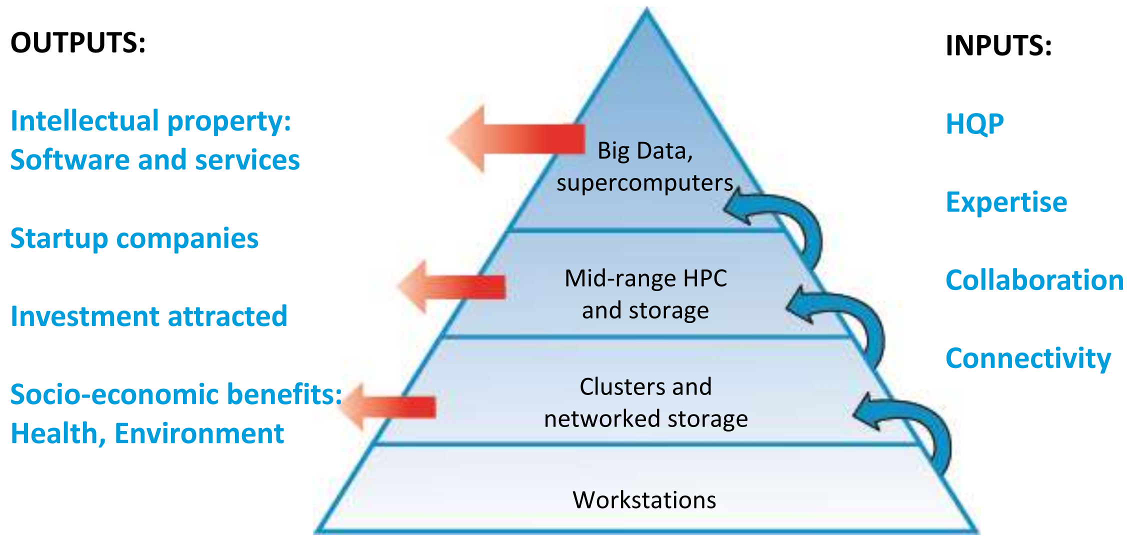

Figure 1: The Advanced Computing Ecosystem

Figure 1 illustrates how researchers use AC and how that generates benefits/outputs. There are fewer researchers at the top of the pyramid, but they are generally highly visible, influential and often part of large international research projects.[8] Of the research groups interviewed, only about 10% were users or potential users of big data or supercomputers with tens of thousands of processors. Many, if not most, researchers and their institutions aspire to ascend the pyramid—tackling larger, leading-edge research problems with more funding. The research ecosystem is dependent on HQP as well as infrastructure, as both are needed to facilitate a researcher's ability to effectively use AC at their current level, and to assist in migrating researchers to larger platforms.

The outputs of the ecosystem are interesting too. Larger AC projects both generate HQP and use HQP, resulting in a feedback loop:

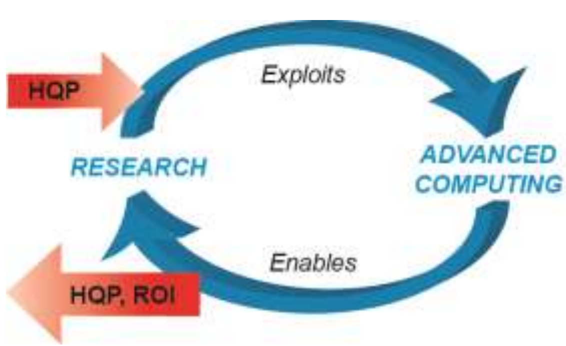

Figure 2: The Advanced Computing and Research Feedback Loop

[8] For example, see http://www.hpcwire.com/2014/01/02/top-supercomputing-discoveries-2013

Accessing Advanced Computing – How Difficult Is It and Why?

Interviews revealed that insufficient access to AC infrastructure and services limits or constrains most researchers, and in many cases is a significant impediment. That leads to the obvious questions of where the researchers' AC support comes from, and whether the problem is a lack of AC resources or merely a researcher's inability to access them.

Advanced computing infrastructure: Local/dedicated versus shared

The data shows that about 50% of research groups rely on dedicated local AC infrastructure (see Table 3), although often not exclusively. This infrastructure ranges from small clusters acquired with departmental resources, to large clusters funded by research agencies or institutions.

Small computing clusters that are locally managed by part-time IT support are typically ineffective for a variety of reasons, including:

- Inefficient use of infrastructure (that is, low utilization of hardware); and
- Inefficient support, typically provided by graduate students or departmental IT staff.

Efficiency is rarely a driving force in academia, and many researchers prefer to have their own small, dedicated cluster with local support, despite the disadvantages. Local support is often perceived as being more responsive, and local HQP are highly valued.

National research funding agencies have recognized the inefficiency of local clusters, and the Canadian Foundation for Innovation (CFI) has mandated that the AC infrastructure it funds should be managed and supported by Compute Canada. This model is known as the "contributed cluster" model, in which research groups acquire their own computing clusters, which are managed and supported by Compute Canada. In exchange for those support services, when the infrastructure is not in use by the funded researchers, Compute Canada gains access to the system for other researchers. The contributed cluster model for AC support has advantages and has worked well in some situations, but can be problematic for two key reasons:

- Each additional contributed AC platform requires additional support staff and services. It is easier to manage one large computing cluster than many small clusters, and a better quality of service can be delivered on a few larger clusters (for example, backup, security and reliability).

- Compute Canada's expertise and focus is on supporting batch HPC computing, which may not meet the needs (or perceived needs) of researchers. Researchers interviewed for this report said they appreciated the support of Compute Canada staff, but sometimes wanted or needed more flexibility.

Cloud computing offers an emerging and highly successful information technology that can be an easier alternative for researchers who believe their most viable option is to acquire their own clusters or rely on batch HPC computing. Cloud computing would instead allow researchers to have their "own" cluster (of virtual machines and storage) hosted on shared infrastructure. However, research use of cloud computing can pose some challenges, which will be explored further below.

Advanced computing support and highly qualified people

The research groups interviewed indicated that they received significant HQP support from the following sources: local (70%), departmental/institutional (35%) or Compute Canada (35%). The percentages add up to more than 100% because many researchers obtain expertise from multiple sources. The percentages are also somewhat elastic, as the definition of where staff and HQP reside is often blurred (for example, the funding for support staff and their reporting lines can come from several sources). Nonetheless, the high reliance on local support correlates with the use of local or departmental clusters for AC.

It is important to remember that the sample was limited, so the percentages cannot be extrapolated to a broader population of research groups, and the classification is somewhat ambiguous (how much access or support is considered "significant"?). Nevertheless, informal discussions with a much wider community of researchers support the conclusions, needs and trends identified in this report.

The percentage of researchers nationally who use Compute Canada can be estimated another way: from the total number of Compute Canada user accounts divided by an estimate of Canadian researchers. That yields a figure of less than 5%. In general, Compute Canada support is focused on experienced traditional users of AC (that is, HPC users) who are at the higher levels of the pyramid (see Figure 1).

Need 2: Institutions and the province need to develop strategies to fund, develop and provide more effective AC support to allow researchers and their institutions to be more productive.

As noted earlier, in Ontario there are about 60 Compute Canada staff at the three major HPC centres and other affiliated centres (such as the Hospital for Sick Children in Toronto), and they are concentrated in a few locations. By contrast, the largest HPC centres in US universities, such as the National Center for Supercomputing Applications (NCSA), may have as many as 200 staff at a single university site.

While there is a gradual trend toward shared AC services, it is apparent that most researchers still rely on clusters and workstations either supported by their department, faculty (this occasionally includes support from one of the HPC centres) or themselves. These researchers are nearer to the bottom of the pyramid in Figure 1. Researchers clearly recognize that there are not enough HQP support staff available to help researchers ascend the pyramid.

The need for HQP has many dimensions, including:

- Expert AC analysts and software support (for example, bioinformaticians)
- Parallel programming and scripting specialists
- Systems support staff (maintaining clusters, data management and so on)

While some research teams indicated that they had sufficient local support, many researchers interviewed had to locate AC support by themselves or with help and advice from other researchers in their discipline or department. About 35% of the research groups reported that there was a significant shortage of access to HQP. This is qualitatively no different than what was reported in 2009 in *Cyberinfrastructure and the Research Process in Canada*. The survey conducted for this ACTION report identified many cases of researchers who were unaware of AC resources and support that were available, either at their institution or provincially. Significantly, almost half of the researchers surveyed explicitly stated that better access to information about available support and services would be of great value to them.

Need 3: There is an urgent and unmet need for better communication about existing AC resources, including training, support and infrastructure.

The three HPC centres in Ontario have great technical expertise and staff. However, they do not have enough resources for marketing and outreach, either through online or more traditional marketing platforms. It is thus often up to researchers to locate the specialist help or resources they need, even as the dispersion of researchers and support in Ontario makes it difficult for researchers and industry to access the existing support and infrastructure.

Advanced Computing Support and Services – Collaboration and Best Practices

Connectivity is an essential element of AC services and support for research at all levels, from networking and shared data to shared information, expertise and support. Sharing and connectivity go hand in hand. A natural hierarchy of infrastructure and expertise exists, with local expertise and infrastructure acting as a gateway to more specialized, shared, and remote infrastructure and expertise. Only a few well-endowed research groups can be locally self-sufficient in terms of AC infrastructure and expertise. Thus, most researchers rely on an informal network of acquaintances, expertise and support to access the resources they need.

Connectivity to HQP expertise varies greatly across and within research institutions. Interviews revealed that some institutions were better able to support their researchers' access to AC than others. According to the researchers interviewed in this survey, these institutions notably included OICR, McMaster University and the University of Waterloo.

Researchers at OICR rate their institutional support highly due to both its resources and its dedicated focus. Few other bioinformatics research groups in Ontario are as well supported as those at OICR.

McMaster University is also a standout because it has an institutional AC support team, which is partly funded by McMaster at an institutional level. The support team at McMaster University is generally highly regarded by researchers; it provides desktop and server administration support, and it assists users in developing grant proposals:

> *The Research & High-Performance Computing Support group provides computing support to the research and high-performance computing communities at McMaster. Our services include desktop and server system administration, web application programming, data visualization programming, data analysis programming, database design, personnel management, and almost any other kind of computer support you might need to support your research endeavours.*[9]

Several faculties at the University of Waterloo also receive AC assistance from IT support groups of about six staff. The funding model used in this instance is a subsidized charge-back.

---

9 http://www.rhpcs.mcmaster.ca/

Aside from institutional support for connecting researchers to AC support, provincial and national initiatives are also highly important. These range from provincial collaboration on AC infrastructure and services (such as the creation of Compute Ontario) to specific cross-institutional collaborative initiatives and platforms, such as Brain-CODE and HPC4Health.

HQP, Data, Computing and Software

As previously noted, almost every survey respondent was dependent on AC—whether that involved access to computing, data storage and management, or software and HQP support. However, the AC needs of every research team and project are different.

Some researchers rely on access to remote data sets and analysis (for example, the archives of the New York *Times* or data from Statistics Canada), whereas others are mostly compute-bound/constrained using community or open-source codes. Others require complex, custom-networked clusters, and still others are reliant on commercial software licenses and packages, as well as the expertise required to install and manage them.

That leads to the following questions: Where are the greatest needs for AC services and support, why and what are the trends? Again, it is important to note that each research team's needs are unique, and this survey was relatively small in scale. However, the survey results are consistent with other reports, such as the 2012 report *Research Computing in Ontario – A Framework for the Next Generation*.

Highly qualified personnel

There is a chronic lack of accessible HQP. Many of the research groups interviewed commented on this, although their needs for HQP vary greatly. In some cases, it is simply that the researchers were unable to connect with existing expertise effectively. There are many types of expertise (for example, bioinformaticians, analysts, parallel programmers and data management experts), and there are many approaches to making better connections and growing the HQP base, such as online resources, internships, mentoring, training and dedicated project support. The situation is particularly problematic in Ontario, which has so many research universities and institutions with hidden pockets of expertise distributed in so many places. There is also the generic problem that researchers traditionally do not look far beyond their own department or research colleagues for assistance.

The long-term trends in research point to a rising demand for HQP for many reasons, including:

- *Bioinformatics and new applications for AC*

  The growth in bioinformatics has exploded over the past decade, with rapid advances in science, biomedical instrumentation (for example, sequencers and biomedical imaging) and declining costs of computation. That has led to a rapid growth in demand for AC specialists to support researchers in rapidly expanding and increasingly specialized disciplines, including genomics, proteomics, metabolomics and systems biology.

  Many other research disciplines have also grown in the breadth and depth of their use of AC, in research areas ranging from remote sensing or sensor networks to virtual engineering and digital humanities.

- *Increasing complexity and sophistication of computer hardware, operating systems and applications*

  In recent decades there has been a rapid increase in the complexity of computer hardware, such as complex multi-core architectures with attached specialized processors and tens of thousands of cores in a single system. These require highly skilled and specialized HQP to extract the maximum performance from such systems and scale up computer models. Likewise, operating systems and job scheduling have become more complex and reliant on sophisticated back-end databases and monitors—all requiring new HQP specialist skills. Finally, modern research applications are web based, and often rely on sophisticated data analysis and visualization tools and applications. This has created an increasing demand for yet another set of specialized HQP skills.

- *Growth in demand for analysts*

  An AC analyst is an HQP with discipline-specific knowledge (for example, in biochemistry) who can help researchers make appropriate choices in AC technology, provide training, and customize AC software and packages to meet the needs of a research team. As noted above, modern AC technology is too complex and diverse for most researchers to be able to understand in detail.

None of these trends is likely to disappear. In fact, the demand for HQP is accelerating.

Data storage, access and management

Data is important to approximately 75% of the research groups interviewed, and about 25% of research groups are currently constrained by data storage, access and management.

- *Insufficient storage*

  This is generally not an issue except for some research groups in biomedicine. (For example, storing and archiving raw sequencing data for cancer mutations requires extensive storage space.)

- *Sharing data*

  Several researchers reported that sharing data was either a problem or an upcoming need. Modest amounts of data can be shared using utilities such as Google Drive or Dropbox, but the limitation soon becomes the network transfer speeds for files and data collections in excess of one gigabyte. Several researchers reported simply shipping hard drives as the most cost-effective approach, despite the unreliability of this method.

- *Curation – managing, organizing and archiving research data*

  This is a common, significant and complex problem related to the management, access and sharing of data, particularly in the long term. Security and privacy are also important issues relating to data curation. It has been recognized at a national level for several years, and national research agencies require that research data generated by grants be archived. There are HQP issues bound up with data storage and management as well. In many cases, researchers reported having created their own ad hoc solutions to storage-management issues, simply because they did not have access to expertise or resources to assist them. The high percentage (70%)[10] of researchers who cited an increasing need for data management reflects their recognition its importance.

[10] See Table 2.

- *Security and privacy*

  About 40% of the research groups interviewed had to deal with data security and privacy issues now or expected they would have to do so in the near future. It is important to note that security and privacy issues are not limited to health data. Researchers across many disciplines and institutions (including engineering, economics and information systems) have to securely access and manage data owned by others. The resulting issues included negotiating legal agreements with the data owners, accessing remote data securely, and ensuring secure storage meets ISO or other standards. Once again, access to HQP and expertise was the biggest impediment for researchers, although some large research groups had well-managed and well-resourced facilities for managing security and privacy issues.

In some cases, local network connectivity is bound up with data access problems, although the baseline access to high-speed data networks in Ontario is generally adequate for most institutions at the present time. However, the increased amount of data that is being anticipated will certainly drive the need for even more bandwidth in the not-too-distant future.

Computing

Several researchers reported that their research was constrained because of their inability to access sufficient computing cycles, making their research less internationally competitive. Globally, the use of 10,000 to 100,000 HPC cores is seen as cutting-edge. In some disciplines, such as climate modelling, open-source software is now able to scale to this many cores. Internationally, leading-edge researchers are running ensembles with thousands of models, effectively using 10,000 cores or more at once.

By contrast, the largest jobs that can generally be run by Compute Canada are at the scale of 1,000 cores. Faced with this problem, the researchers interviewed reported having to resort to a number of coping strategies, such as reducing the scope of their research to problems that could be solved using fewer cores, or by building collaborative relationships with international researchers who could provide access to systems with more cores (such as those in Washington state or Zurich, Switzerland). However, accessing international facilities comes with many additional constraints: generally, only principal investigators are given accounts (excluding graduate students), and the project and author lead titles then fall to the researchers at the international site hosting the HPC, instead of the Canadian collaborators.

In addition, the lack of competitive computing access often results in Canadian researchers' publications being poorly recognized. That, in turn, makes it harder for Canadian researchers to attract and retain graduate and post-doctoral students in an international market, a problem which is further exacerbated by the fact that universities in Ontario (and the rest of Canada) are not graduating as many HQP in AC as are needed.

Traditional HPC is batch oriented (batch HPC) and requires remote access/login to Linux-based supercomputers. This is the primary model of AC that Compute Canada supports. This survey found that about 35% of researchers used this service, but a larger percentage of researchers used their own clusters and workstations (local or interactive HPC) in combination with or in preference to batch HPC. This preference is based on a mix of characteristics:

- Local control of scheduling, installed software and hardware, custom graphics processing units (GPUs) or solid-state disks (SSD).
- The need for greater interactivity, which gives researchers the ability to view and interact with running applications on the fly.

Most of the limitations of batch HPC can be overcome by using a variety of techniques, but many researchers are reluctant to invest the time and resources to do so.

Another way to access HPC is through custom-built web applications that act as front-ends or workflow managers for batch HPC. These are common and popular in bioinformatics; Ontario research groups, including OICR and OBI, are leading the use and development of these platforms. Yet there is still a need for more interactive and locally controlled access to HPC. Cloud computing can and does provide this, and groups such as SOSCIP are providing access for select projects. A viable alternative to traditional batch HPC could include using cloud computing to resolve the problem of researchers having to acquire and maintain their own cluster and workstation hardware, while providing local control and interactive virtual machines.

Cloud computing for research

The depth and breadth of tools, technology and hardware support available in cloud computing is rapidly evolving, so any assessment of its importance for research needs is inevitably a moving target. In the previous decade, the focus of making HPC easier to use was grid computing: connecting HPC computers into a seamless network with features such as a single login and automated remote file and data transfers between systems. By the late 2000s, cloud computing was emerging as a viable commercial technology and the HPC community was starting to take interest. The earliest major studies on using cloud computing for HPC started in 2009 and produced mixed results.[11] Some of the key findings in this NSF multi-year study included:

- *Scientific applications with minimal communication and I/O [input/output] are best suited for clouds.*
- *Clouds require significant programming and system administration support.*
- *Public clouds can be more expensive than in-house large systems. Many of the cost benefits from clouds result from the increased consolidation and higher average utilization. (Trader, 2012)*

Since these findings in 2009, cloud technology and software has evolved considerably. The only finding that remains largely unchanged is the last one: Public clouds are not necessarily cheaper than large in-house systems. That finding has been reinforced in later studies conducted by groups such as Compute Canada. It should be remembered that these cost comparisons can be deceptive. Often academic institutions don't include staff support costs, heating/cooling, power and back-up systems in the total cost comparisons, as they are taken from a different departmental budget. Industry implementations certainly show that a cost savings can be realized when cloud resources are a feasible replacement for smaller, dedicated clusters and in consolidating smaller clusters and systems into a single virtualized data centre.

Thus, cost comparisons can be very deceptive, as public clouds often offer a broader range of services than traditional academic HPC, such as 24/7 support and failover/restart (for example, Hadoop clusters). As researchers begin to demand higher quality resources (whether computing, data and/or support), the costs to deliver to the expected "commercial grade" increases proportionally. These levels of service are outside the normal delivery of academic IT infrastructure and environments.

Recent technology advances are significantly affecting the other two findings. First, the communication and input/output (I/O) overhead or costs of cloud services are being significantly reduced. Cloud computing relies on virtualizing the underlying hardware. A small kernel program, called a hypervisor, allows multiple operating systems to run transparently on the same hardware platform. The hypervisor emulates bare hardware, such as memory and I/O channels. Significant improvements in the performance of this emulation are being driven by the wide commercial adoption of virtualization and cloud computing. Nevertheless, applications that are communication intensive (especially if they use specialized HPC communication hardware and drivers such as Infiniband) will continue to be better suited to traditional HPC. Such

[11] See http://archive.hpcwire.com/hpccloud/2012-02-01/learning_from_clouds_past:_a_look_back_at_magellan.html

applications make up only a fraction of the research community's use of AC. From both a technical and an administrative viewpoint, it is arguably better to move the computing jobs that do not need specialized HPC hardware off of Compute Canada and onto (public or private) commodity cloud systems.

The second finding of the study cited above relates to the programming and systems administration support costs of cloud computing. In recent years there have been significant improvements in the reliability, ease of use and installation of cloud computing—including both commercial cloud platforms (such as Amazon, Microsoft and VMware cloud platforms and hosting) and open-source cloud platforms (such as OpenStack and Eucalyptus). Most popular cloud platforms now have point-and-click web interfaces that allow users with limited IT experience to create cloud instances, monitor their performance and share data.

Another relevant issue with cloud computing in Ontario is data security and privacy, especially as it relates to legal jurisdiction of data storage. For some research purposes, data must be stored securely, and it is preferable or even legally mandated that the data be hosted in Ontario and/or by an Ontario company or organization.

Finally, other relevant issues with cloud computing involve the cost models for data storage and network traffic (the NSF study cited above was for a single-site cloud installation on academic research networks). There can be hidden costs and performance issues in cloud computing that a locally based cloud solution can overcome, although "first mile"[12] issues (a researcher's local connection to ORION's backbone in Ontario) are often more problematic than provincial or national network performance.

Software support

Each research group's use of AC is unique and relies on a suite of domain- and research-specific software packages built specifically to support that research in terms of data analysis and manipulation, modelling and simulation, and visualization. Acquiring, configuring, installing and maintaining such software package suites generally requires the ongoing support of HQP (whether they are local graduate students, institutional support or Compute Canada staff).

Aside from the difficulty of accessing HQP for software, a further complication is negotiating licenses for commercial software packages. A significant proportion of researchers interviewed relied on commercial software—especially in engineering disciplines, where a majority of researchers interviewed rely on

[12] The term "last mile" is often used in industry instead of the equivalent term "first mile," but the term "first mile" is more user/service focused.

commercial software packages for computational fluid dynamics (CFD), finite element analysis (FEA) and software development (especially MATLAB) from a plethora of vendors. Many institutions and researchers already have such licenses, although negotiating changes to licenses (such as upgrades, new installations and floating licenses) is tedious and difficult for researchers. Several of the researchers interviewed commented on the difficulty these negotiations present.

The recommendations that follow address how these needs might be better supported at a provincial level.

Networking and Connectivity

Broadband networks underpin everything related to AC. Modern society is utterly dependent on high-speed networks, from smartphones and personal videoconferencing to video-on-demand. The research community's reliance upon networking to support AC reflects these global trends. High reliability and high-speed networking is often taken for granted—until it fails.

Governments have long recognized the importance of high-speed networks to support R&D. The Internet itself was founded as a US government–funded project to support R&D. In Canada, the federal government has funded the national network of CANARIE since 1993.[13]

At a provincial level, optical regional advanced networks (ORANs), such as ORION in Ontario, connect local research and educational institutes. The role of these research network providers continues to expand as they invest in providing researchers and institutions with more network-enabled shared services. This trend is happening internationally, nationally and, specifically in Ontario, through ORION's Nebula Cloud Services (www.orion.on.ca/cloud-services).

The importance of connectivity and ongoing investment in wireless and optical networks is underpinned by the growth in increased data traffic. On provincial, national and international levels, the annual growth in network traffic has been on a long-term trend of about 50% per annum, but that is expected to increase significantly with the deluge of data that is starting to occur as a result of even larger research datasets, the "Internet of Things," and the use of mobile devices in research.

Most research groups interviewed were satisfied with the network connectivity that they had. In some cases there were issues of connectivity within an institution. Network providers such as ORION generally have only a single Point of Presence (PoP) or connection point per campus; it is then up to the institution or researchers to connect their local computing systems by using their institutions' high-speed campus network.

Another issue raised by one researcher was network latency/reliability leading to software failure. This can be complex to diagnose and fix, as it involves interplay between the application and local and provincial network connectivity. In general, network latency is not an issue for most researchers or modern networks, and commodity cloud applications such as videoconferencing are continually improving and upgrading.

[13] http://www.canarie.ca/en/about/ataglance

Finally, some research groups reported using inefficient and unsafe practices, such as shipping hard disk drives instead of using networks or cloud services. Storage capacity is generally increasing by a factor of 10 every five years. That means that it is increasingly impractical to copy entire large research databases (more than one petabyte) over the Internet. One solution is to co-locate the application with the data. Yet sharing and collaboration will still drive the need to transfer this data over networks. As such, it becomes even more important to have dedicated research networks with distributed connectivity, such as ORION and CANARIE.

The ACTION Path Forward – Meeting Researchers' Advanced Computing Needs

The previous sections of this report analyzed both the depth and breadth of researchers' needs for AC support. Three key generic needs were identified:

> Need 1: Exploiting AC requires HQP, and access to HQP limits most researchers in Ontario more than the difficulty experienced in accessing HPC infrastructure.
>
> Need 2: The province needs to develop new strategies to fund, develop and provide more effective AC support to allow researchers and their institutions to become more productive.
>
> Need 3: There is an urgent and unmet need for better communication about existing AC resources, including the available training, support and infrastructure.

This section of the report builds on the conclusions above to identify specific recommendations to provide better AC support for researchers in Ontario. Several assumptions are made in the recommendations below. First, this report assumes no significant short-term changes in national funding and national AC organizations, such as Compute Canada. Second, these recommendations are focused on what can and should be done at a provincial level, rather than at an institutional level.

This survey indicated that researchers need local AC support (instead of just provincial or national support), especially for face-to-face mentoring and supporting the majority of researchers in Ontario who:

- are not traditional users of HPC (for example, through Compute Canada); and/or
- are not part of self-sufficient, larger research teams; and/or
- require access to AC support which is not provided by the resources at Compute Canada.

Given the importance of AC, it is in the best interests of research institutions to reflect on their current institutional support of AC. However, it is not in the scope of this report to make recommendations on institutional support.

Finally, it is not within the scope of this report to make recommendations on funding, investment or organizational plans for implementing the recommendations below. That has been done in reports such as the

Gulyas report (see Appendix D).

The following recommendations have been developed based on the needs identified above.

Recommendation 1: A program to develop expertise

The shortage of HQP and the importance of increasing the numbers of HQP available for AC support has been well documented in both this study and other reports. There are many ways to develop HQP, including:

- *Training courses*

  Training opportunities range from lectures and short courses to degree programs. The three HPC centres in Ontario already provide these to varying degrees, although they are neither widely publicized nor coordinated on a provincial level.

- *Internships and job placement programs*

  Several researchers commented that they have funds for interns or student support, but find it difficult to locate students with the right expertise or availability. Some institutions have programs and projects that connect IT expertise and students (typically in computer science and IT departments) with researchers in other departments that need access to such expertise. There are also some cross-institutional programs in place (for example, between OICR and the University of Waterloo, as well as OCE's TalentEdge program). More could be done to foster and promote such initiatives at the provincial level.

- *Conferences and workshops*

  The annual Canadian conference of AC/HPC is the High Performance Computing Symposium (HPCS). This conference has traditionally focused on systems support staff rather than researchers. Local Ontario conferences, workshops and research days could complement HPCS and provide additional avenues for researchers to connect with AC expertise provincially. This has already commenced with the Compute Ontario research day hosted by the Perimeter Institute in May 2014, and should continue with other events in the future.

- *Online resources*

  A great deal of material for training staff and students already exists online, although it is not often easy to find relevant and up-to-date material that has been reviewed. There is opportunity here for Ontario to collaborate with international efforts to better develop these resources. There is also a clear overlap and synergy between Recommendation 3 (an Ontario AC collaboration portal, below) and this recommendation.

Recommendation 2: Better access to cloud computing

As discussed above, current AC support is not effective for many Ontario researchers, and there is a need to develop and provide more effective AC support to researchers. Cloud computing, with its self-service, on-demand, as-needed access to AC, offers a technology platform that can have a significant impact in meeting this need.

In Ontario there are several cloud projects in the research and education sector, including:

- A pilot cloud project for researcher access at the University of Western Ontario, as part of the SOSCIP project
- Cloud research projects at institutions including the University of Toronto and York University, which are extending the capabilities of current cloud platforms
- Experimental installations of cloud software at several research institutions, including OICR and CMC Microsystems
- Pilot cloud systems for support of education at Carleton University and the University of Ottawa

In the education sector, cloud resources may be able to replace the traditional computer labs, which are widespread across university and college campuses. Such labs are costly to build, maintain and operate, and are relatively underused. By contrast, accessing classroom workstations in the cloud can be much more cost-effective and provide a better level of service (such as 24/7 access to workstations from any computer connected to the Internet). There is a clear precedent for this in the North Carolina VCL[14] cloud, which has been operating successfully and growing for almost a decade.

Several researchers related how they commonly buy small clusters or high-end workstations (worth $5,000–$10,000) when their research needs outstrip the capacity of typical desktops and laptops, relying on graduate

[14] See http://vcl.ncsu.edu

student support for their local cluster. In some cases, these departmental clusters can be supported by local IT support staff, but it still may be more cost-effective and foster better service delivery to use cloud-computing services rather than purchasing a small cluster. The ACTION workshop also revealed significant research interest in cloud computing (in fact, several research groups have already been experimenting with commercial cloud services, including OICR).

Many researchers may be better served by cloud services than by purchasing their own small clusters for a variety of reasons, including:

- *Cost effectiveness*

  Typical costs of cloud computing are in the range of 10 cents per core-hour, or about $800 per core-year. That is low compared to the cost of buying a workstation cluster, assuming that the workstation cluster is not being run flat out (that is, 24/7 computing).

- *Flexibility and scalability*

  Cloud services are pay-as-you-go and can generally be customized for the resources required. Importantly, a researcher can also utilize on-demand scale-up for the amount of computation, memory or storage required, which is not practical for a local in-house cluster.

- *Better support and maintenance*

  A cluster or workstation needs backup, support and maintenance. This is often neglected in departmental clusters, or is merely a waste of time and resources.

- *Simplified sharing and remote access*

  A local workstation or cluster is generally not readily accessible remotely, whereas cloud resources are remotely accessible and shareable.

Some research groups have been experimenting with commercial cloud providers such as Amazon Web Services, but that has not been widespread to date. Cloud services have yet to make significant inroads into research computing in Ontario for a variety of reasons.

The negatives associated with cloud services include:

- *Complexity of setting up a cloud-computing environment*

  Setting up a cloud-computing environment requires IT skill and expertise. Most researchers have little or no familiarity with cloud services, nor do they have any interest in exploring such services. They often find it simpler to resort to more familiar methods, such as buying a physical cluster.

  However, there is a rapid evolution in online, web-based dashboards for interactive setup and configuration of cloud computing (for example, Amazon and Azure APIs, and open-source equivalents for OpenStack[15]). These make it possible for researchers with limited expertise to set up and deploy cloud services.

- *Hidden costs of cloud computing*

  While computing costs for cloud resources can be surprisingly low, there are often hidden costs, such as network traffic costs or costs for backup/redundancy. ORION's network will be key in reducing network costs, offering high network throughput and faster response times for Ontario researchers.

- *Software licensing*

  Commercial licenses for software can preclude hosting the software in the cloud, or complications may be created in dealing with remote license servers. This drawback is gradually decreasing with the introduction of cloud licensing by major software vendors.

- *Data access and ownership*

  Most public-access commercial cloud offerings are hosted in the US. This creates privacy issues and may violate contractual security agreements for data hosting. Ontario-based cloud resources could solve this problem, provided that appropriate data hosting, security, privacy agreements and facilities are in place.

- *Network bandwidth*

  Network bandwidth can be an impediment to remote desktop access to cloud computing. As noted above, ORION's network offers a solution for Ontario's researchers and cloud providers.

[15] See https://www.openstack.org/software/openstack-dashboard/

- *Lack of local support*

  Commercial cloud vendors offer web and phone support for knowledgeable users, but it is generally remote and targeted at IT professionals. Most researchers benefit from experience, and in some cases they rely on local HQP support.

- *Capital expenditures versus operational expenditures*

  A move to cloud computing shifts funding from capital expenditure on infrastructure to operational expenditure for services/cycles. While this is usually attractive, it can be problematic for traditional grant-funding models that are designed to fund infrastructure, not services. Research grants do not often allow flexibility (such as in shifting funding from capital expenditures to operational expenditures).

The benefits and drawbacks for using cloud services will be different for each research project and community, and will shift as technology evolves and cloud resources mature. However, there is obviously a significant research community that would benefit from using cloud services as an alternative to their current computing facilities. Given the need and the productivity potential for better access to cloud computing for researchers, the question is: How should cloud computing be provided and supported, and by whom? Potential sources include commercial cloud providers in Ontario, the HPC centres, a consortium of providers or network-centric organizations like ORION.

Recommendation 3: An online community connecting researchers with advanced computing resources

There is always a need for more AC infrastructure. However, the evidence suggests that better access to information about current AC resources and services is a far more pressing need. Researchers are consistently unaware of available AC services and facilities, which other researchers in the province have similar needs and with whom they might collaborate. Accessing such information is currently extremely difficult. While the three HPC centres in Ontario (HPCVL, SciNet and SHARCNET) have websites, they are not always comprehensive or easy to search, and many researchers are not even aware that these centres exist.

Developing a community information and collaboration site is technically quite easy, but it would be equally easy to devote a large amount of effort to developing a website that is difficult to use and to maintain. Therefore, several principles must be considered when developing such an online community:

- *Usability*

  The online destination must be focused on meeting the needs of researchers, with a design and user interface that are engaging, easy to navigate and that leverage modern web technology.

- *Sustainability/critical mass*

  The site should be self-sustaining, ideally by making it easy and engaging for researchers and AC service providers to use and by delivering value, thus providing them with incentives to keep their information up-to-date and participate in relevant forums and information feeds.

This Ontario AC website should have a clear focus on, and input from, AC research users, service providers and stakeholders. It would need to support the development of special interest groups and communities, and leverage existing datasets and social media websites, without reinventing or duplicating existing data.

Recommendation 4: Collaboration on privacy and security template agreements and facilities

Privacy and security of information are of widespread concern across research and industrial communities. Although this is a complex landscape, collaboration is already occurring—and is even increasing. However, many researchers are unaware of the capabilities and expertise that currently exist in organizations such as HPCVL, OBI, OICR and HPC4Health. Could closer provincial collaboration assist those groups? If so, what form would that collaboration take? Privacy and security could be a special-interest focus of a provincial AC collaboration site.

Recommendation 5: Provincial collaboration on platform development, especially bioplatforms

Considerable effort is being expended provincially, nationally and internationally on developing software platforms to support research (for example, CANARIE's research middleware program,[16] or NSERC's Science Gateways program[17]). Unfortunately, information about platforms—specifically, their availability and capabilities—is often not readily accessible. This leads to repeated re-invention and redevelopment of new

[16] http://canarie.ca/en/middleware
[17] http://www.nersc.gov/users/science-gateways/

platforms, rather than reusing and leveraging existing platforms and their components.

The biomedical research community in particular is rapidly evolving and getting involved ever more closely with clinical data, clinicians and personal health information collection and management. This is increasingly dependent upon complex, evolving federated IT systems and platforms, which are often being developed in islands of expertise. Closer collaboration and awareness of platforms, and sharing expertise on a provincial level, could be valuable.

A specific example that provokes this recommendation involves the researchers at Robarts Research Institute, who need to share biomedical imaging data and tools with collaborators. Instead of shipping data to collaborators on hard discs, the data and tools could be securely shared using provincial cloud computing.

Once again, bioplatforms could be a special-interest focus of a provincial AC collaboration site. The forms that such collaboration could take include:

- Better information about available platforms and resources, such as the capabilities of the platforms, their architecture, and how to download/install or access them remotely
- Better interoperability between platforms, as well as the sharing of developer expertise
- Joint development efforts to broaden the base of platform support and development

Recommendation 6: Research data management and archiving

Research data management is becoming increasingly important, and a majority of researchers surveyed considered this either a significant need or an emerging need. Many research communities have developed solutions, although others struggle with how to manage their data or archive it. A long-term, generic solution for managing research data is extremely complex and has been the subject of multiple national workshops and projects, for example, through the Canadian University Council of Chief Information Officers (CUCCIO), the Canadian Association of Research Libraries (CARL), Research Data Canada (RDC) and the Scholars Portal program at the Ontario Council of University Libraries.[18] Better provincial collaboration on expertise, facilities, tools and best practices would benefit many research communities.

Other jurisdictions have funded major efforts to support research data management, such as the Australian

[18] http://ocul.on.ca/node/135

National Data Service,[19] which is an ongoing program for national research data management in Australia; it was formed in 2009 with more than AU$50 million in funding.

It should be emphasized that research data and its management is an asset, not an overhead. In the future, the organizations that have the best online research data collections and management will be more competitive for funding and will generate better socio-economic outcomes. There are already several examples of Ontario research institutions that have internationally regarded research data collections.

The need for research data management support overlaps the need for security and privacy. The scope and stakeholders in research data management are very broad, ranging from libraries to research hospitals. Any next steps taken at the provincial level to manage research data must engage all of these stakeholders. Given the difficulty in developing general strategies and tools for research data management, and the exorbitant costs of such efforts in other jurisdictions, a more suitable approach for Ontario could include providing researchers with better tools and resources for data management and archiving research data as a first step. Some cloud service providers have facilities for automating archiving of research data and this may be an attractive and inexpensive option for researchers.

Recommendation 7: Better industry access to advanced computing

Industry is not well served by AC access in Ontario, with the exception of large companies that can afford their own HPC/AC. The need for better industry access to AC, especially for small- to medium-sized businesses, has been globally recognized, and there is a wide range of projects and services in place to provide this (for example, SOSCIP in Ontario, the CANARIE DAIR project[20] and the UberCloud crowd-sourcing model). The collaboration site described above could provide industry with easier access to AC.

There are additional sector-specific opportunities. Ontario has a thriving digital animation (DA) industry sector, with a direct economic impact of $83 million[21] in 2010 and employing about 2,000 people across several hundred companies. The Ontario DA industry is highly diversified and specialized, and it is populated by

[19] http://www.ands.org.au/
[20] http://www.canarie.ca/en/dair-program/about
[21] Nordicity. (March 2012). Economic Profile of the Computer Animation and Visual Effects Industry in Ontario, 2008-2010. Available online at: http://www.omdc.on.ca/Assets/Research/Research+Reports/Economic+Profile+of+the+Computer+Animation+and+Visual+Effects/Economic+Profile+of+the+Computer+Animation+and+Visual+Effects+Industry+in+Ontario_en.pdf

many small- to medium-sized enterprises. The DA industry is also dominated by two key needs: specialist DA staff ("artists") and AC infrastructure. Large DA projects are highly dependent on large investments in computing and storage infrastructure by individual DA companies—investments that are complicated by the cyclical nature of the industry. On-demand access to computing and data servers would make the industry more efficient and competitive: DA companies could access secure specialist DA cycles and services in shared, secure data centres, rather than each DA company having to invest in its own AC infrastructure and support staff. Such initiatives already exist in London, England, and Vancouver, BC. Aside from its economic impact, a DA shared-services cloud network could be a useful precursor or demonstrator for similar specialist shared-service clouds supporting small- or medium-sized businesses in sectors such as engineering and manufacturing.

Conclusion

The demand for AC support and services in the research sector will only continue to grow, with trends such as:

- An increasing focus on large-scale, multi-disciplinary research projects to tackle national and provincial research priorities;
- The growing use of modelling and simulation to reduce the high cost of wet-labs and physical prototypes and testing; and
- The massive growth in data generation and analysis (for example, in biomedical data and sensor networks).

The demand for HQP is outstripping the demand for AC infrastructure, therefore a greater focus on how to better connect researchers to both the available infrastructure and expertise, as well as how to grow that expertise base, is needed. That focus must be supplemented by ongoing investment in AC infrastructure, in order to ensure that Ontario's researchers have access to world-class facilities and the expertise to make the most effective use of that investment.

Appendix A – The ACTION Project and Advanced Computing Organizations in Ontario

Further details about the ACTION project, such as its motivation and presentations, are available on ORION's website at www.orion.on.ca/action.

The ACTION technical working group consists of the following members:

Bill Appelbe	ORION www.orion.on.ca
Shiva Amiri	Ontario Brain Institute www.braininstitute.ca
Allison Barr	Ontario Ministry of Research and Innovation www.ontario.ca/ministry-research-innovation
Mike Bauer	SHARCNET www.sharcnet.ca
Michael Brundo	The Hospital for Sick Kids/HPC4Health www.sickkids.ca
Ken Edgecomb	High Performance Computing Virtual Laboratory www.hpcvl.org
Chris Loken	SciNet www.scinethpc.ca
Lincoln Stein	Ontario Institute of Cancer Research www.oicr.on.ca
Rhonda Tannenbaum	Ontario Genomics Institute www.ontariogenomics.ca
Ron Van Holst	Ontario Centres of Excellence www.oce-ontario.org

Many other people provided feedback and support, and the authors especially wish to thank the many researchers who took time out of their busy schedules to meet and candidly discuss their specific AC needs; the committee members, who supported the project; and ORION, for its financial and organizational support.

For further information or feedback, please contact Bill Appelbe (bill.appelbe@orion.on.ca).

Appendix B – Researcher Survey

The following survey was used during meetings with researchers. It was used primarily as a template discussion document: Answers were recorded in hard copy for each meeting then transcribed into a Google Drive online document. Each research group had unique perspectives and needs that were documented.

Detailed comments were recorded for all issues raised. Meetings were held with about 50 leading research groups across seven Ontario universities and research institutions, including McMaster University, Ontario Institute for Cancer Research (OICR), Queen's University, the University of Toronto, the University of Waterloo, the University of Western Ontario and York University.

Date___________ Attendees____
Research Description ___________________________
Funding Sources __________________________

RESOURCE	CURRENT USE/NEED	FUTURE TREND	COMMENTS
Compute "cycles" – where/how?			
o Access to compute time o More "powerful" or "accessible" compute?			
Software applications access[i] open source, commercial, and/or in-house?			
o Modelling and simulation o Data analysis			
o Visualization			
o Workflows, IDEs[ii]			
Data access:			
o IP, privacy & security barriers			
o Latency – fast data access o More storage			
o Organizing & archiving data o Publishing data			
HQP access – where/how is expertise sourced from?			
o Casual help/support[iii]			
o Ongoing - experts/analysts[iv]			
R&D Project Management[v]			
Key Needs – productivity, sustainability, growth			
Strategic Opportunities			

Figure 4 – Researcher Survey

Appendix C – Detailed Outcomes from Researcher Survey

C.1 Computing

Each research group interviewed has a different research agenda, which results in different computational needs. The diversity of computing needs is quite dramatic and includes:

- Graphics processing unit (GPU)–based computing on workstations and clusters, optimized for biomedical imaging
- Research problems limited by memory access and a lack of large shared-memory clusters
- Cloud-based computing research—investigating new cloud platforms
- Data analysis, which was constrained by disk access; hence the researchers used custom clusters with fast solid-state discs (SSDs)
- Large computational problems that could be satisfied by very large computing runs in the public cloud

Supporting such diversity presents a challenge, but researchers are very creative. The observation that only about 35% of these researchers use Compute Canada facilities correlates with the diversity of computing needs. Compute Canada may not have the resources to support such diversity, and the argument could be made that providing such support is not a part of Compute Canada's mandate. This presents an interesting conundrum, as the Canadian Foundation for Innovation (CFI) has mandated that new computing systems it funds should be managed by Compute Canada.

While the use of new computing platforms is growing, there is still a strong demand for traditional batch HPC. Four research groups commented directly that they had to restrict their research endeavours to the available computation resources, significantly affecting their ability to be internationally competitive (such as by constraining their ability to undertake high-impact research, attract post-doctoral students and so on). Their experience has led them to conclude that computational science publications in high-profile research journals such as *Nature* need to be groundbreaking in the sense of tackling larger and more complex problems than have been previously achieved, which requires very large scale computing resources. At present, Canada has few resources that could be characterized as Tier-1 supercomputing systems.[22] Three research groups commented on the aging computing infrastructure at Compute Canada as being a constraint. The typical

[22] Gulyas, G. (2013). See Appendix D.

lifetime of an HPC system is three years, which can sometimes be extended to five years. Beyond that age, the performance of the systems is often not competitive with what many researchers can acquire with a small custom cluster.

C.2 Data

As discussed earlier, for many research groups research data is as important as computation. Research dependence upon data is forecast to increase, and that dependence includes:

- *Secure data access*

  Secure access to private or confidential data in joint projects including post-secondary institutions, industry and other organizations. Most biomedical groups already have secure data access and privacy agreements in place, although these take time and resources to negotiate and develop. More than 50% of the groups interviewed were either dealing with security and privacy issues now, or expected they would have to deal with them in the next few years.

- *Data management/archiving*

  Managing and archiving data were mentioned as being significant and/or of growing significance for about 25% of the research groups interviewed.

- *Data access, sharing and collaboration*

  These items are often tied together. Four of the research groups interviewed shared data by shipping hard drives, as they believed there were no secure alternatives for sharing gigabytes of data. While fast provincial networks exist, there are often first mile problems connecting researchers to high-speed networks within their campuses. Some research groups have resorted to running cables between their labs.

  Data access often relies on establishing trust and long-term collaboration. In some cases, data access was hampered by the procedures and policies of the data owners. Several researchers commented that access to Statistics Canada's data is not easy, as it relies on physical access to a secure location.

- *Data analysis*

  Data analysis is increasingly important to many researchers in areas such as bioinformatics, economics, science (for example, climate modelling) and public health. About 30% of research groups spent a significant part of their research activity undertaking data analysis.

Most researchers used structured data, with one or two exceptions. Thus for the research groups interviewed there has not yet been a major demand for database tools and technology specifically designed for dealing with large unstructured data sets (for example, Hadoop).

C.3 Highly Qualified Personnel

As noted, improved access to HQP is a major need for many researchers. About 70% of research groups relied on their local network for HQP support: graduate and undergraduate students, post-doctoral students or HQP hired for group or departmental support. Many groups rely on their graduate and undergraduate students to acquire the necessary IT skills, especially in science disciplines outside of computer science. This may involve self-directed learning and training courses. However, this can be very costly, in terms of both time and also the quality of the IT skills that students acquire through self-learning. For example, self-taught IT skills often include programming, but rarely include software maintenance, testing and usability. This means the software projects and products produced by self-taught programmers often only last as long as the project lasts, and may involve unnecessary software development that could have been avoided by better IT choices.

Not surprisingly, in computer science and IT, almost all researchers interviewed had reasonable access to HQP, as their students all possessed an IT background. Outside of the sciences and engineering, however, almost all research groups interviewed relied on institutional support or HQP support from Compute Canada, and that access was very dependent upon the researchers' ability to form connections and locate expertise outside of their departments.

Reliable HQP accessibility is highest in those institutions and research groups that have formed good relations with sources of IT expertise (for example, in computer science departments).

The range and frequency of IT skills that research groups needed or wished they had better access to included:

Job Classification	*Number of research groups needing skill*	*Description*
Systems administration	3	Support of local clusters
Software	3	Installing software packages, tuning

applications support		
Programming	4	Development of new codes, parallelization
Business development	3	Assistance with grant proposals, industry engagement
Analysts/domain specialists	5	For example, bioinformaticians

Table 4 – Distribution of HQP Skill Needs

The list of skills required is both broad and deep. In several cases, researchers had funding but could not find an appropriate person to hire.

Several researchers commented that they could engage in broader research activities and collaboration with industry if they had additional HQP support.

C.4 Software

Researchers interviewed used both open-source[23] and commercially licensed software, although the majority of researchers used open-source software. Commercially licensed software is sometimes preferred over open-source software, either because of its capabilities or the support that is offered, even though in many cases an open-source product exists with capabilities comparable to commercial software, for example, using OpenFoam for computational fluid dynamics (CFD) instead of commercial CFD tools such as Fluent or CFX. Commercially sponsored research may also require or prefer researchers to use commercially licensed software.

Open-source software generally requires that the user install and maintain the software and all the packages it depends on. That often requires specialist HQP support. The Ontario HPC centres all provide support for accessing, installing and maintaining open-source software on their systems; although generally not for users who want to install software packages on their workstations or clusters. Several groups are developing new open-source products, acting as lead support on products or contributing to open-source projects. This should be encouraged and supported at a provincial level by indirect means such as encouraging greater collaboration

[23] The term “open source” is used loosely above, to mean software which is available free of charge and can be modified and redistributed. Technically, open-source software also includes commercially licensed software (provided that the source code is made available).

on software platform design and re-use. Authoring and supporting open-source packages that are widely used generates both significant recognition and potential revenue from spin-offs that provide commercial support for such software.

Several groups used visualization software and packages, but generally found that the local support was good.

Acquiring and installing software packages, especially commercial software packages, was a problem or issue for at least three of the research groups interviewed. This is especially true in engineering disciplines, where open-source software is less common than in the traditional sciences. A directory of which packages are installed where, and more provincial collaboration on licensing, would assist many researchers.

C.5 Broadband Networking and Connectivity

As noted in the section entitled "Networking and Connectivity" (page 36), high-speed connectivity underpins AC and is often taken for granted. The specific needs of researchers include:

- Strong, reliable campus connectivity. This is the "first-mile/last-mile" problem. Generally, network bandwidth and latency are sufficient going in to the campus (such as ORION's network and PoP equipment), but it is the connection from the PoP to the campus that can cause a bottleneck. Each stage of a connection from that PoP to a local computer system can require relatively costly investments and local network hardware upgrades. It can also be a mix of technologies—optical, Ethernet, wireless and so on. On some campuses, this infrastructure investment is a relatively costly charge-back to the researchers or their department. Research groups who were interviewed identified cases where the researchers simply bypassed campus IT support and ran their own network connections.

- Cloud computing, as an alternative to shipping hard disk drives (HDD), is a pervasive practice by researchers. As noted earlier, several research groups interviewed regularly used HDD shipments as a way to move large files around. The principal reason for this approach was the lack of network bandwidth, although security was also cited. In some cases, the researcher's collaborators were commercial companies who had restrictions on how data could be transferred or shared. In the example of the Robarts Research Institute's researchers, they shipped HDDs with both the data (3-D biomedical images) and the application(s) to view that data. It would be considerably more efficient to put both the data and application in a secure, provincially shared storage/computing facility (that is, a community cloud system).

Appendix D – References

CANARIE; CUCCIO; Compute Canada. (January 2010). Cyberinfrastructure and the Research Process in Canada. Toronto, ON: Findings from the Canada HPC workshop, December 17–18, 2009.

Compute Canada. (March 2014). The Compute Canada strategic plan, 2014–2019. Available online at: https://computecanada.ca/en/about-us/news/general-news/345-compute-canada-calcul-canada-strategic-plan-2014-2019

C3.ca (now Compute Canada). (January 2005). Engines of discovery: The 21st century revolution. Toronto, ON.

Gulyas, G. (2013). Compute Canada – Investment Outlook. Toronto, ON: Ontario Ministry of Research and Innovation, prepared by Greg Gulyas Consulting. (CONFIDENTAL).

High Performance Computing Council. (October 2008). Ontario High Performance Computing Strategy. Toronto, ON: Ontario Research Fund Advisory Board.

HPCwire. (2012). Learning from Clouds Past: A Look Back at Magellan. San Diego, CA. Available online at: http://www.hpcwire.com/2012/02/01/learning_from_clouds_past:_a_look_back_at_magellan/

Industry Canada. (2007). Mobilizing Science and Technology to Canada's Advantage—2007. Ottawa, ON: Industry Canada. Available online at: http://www.ic.gc.ca/eic/site/icgc.nsf/eng/h_00856.html

Industry Canada. (April 2014). Digital Canada 150. Ottawa, ON: Industry Canada. Available online at: http://www.ic.gc.ca/eic/site/028.nsf/eng/home

KPMG. (November 2011). An assessment of future demand and needs of the High Performance Computing system in Ontario. Toronto, ON: KPMG.

Ontario Expert Panel. (June 2012). Research Computing in Ontario – A Framework for the Next Generation. Report submitted to MEDI. (CONFIDENTIAL).

Ontario Ministry of Research and Innovation. (January 2014). Supporting World-Class Research: Ontario Investing in Innovation and Creating Jobs. Toronto, ON: Ontario Ministry of Research and Innovation. Available online at: http://news.ontario.ca/mri/en/2014/01/supporting-world-class-research-1.html