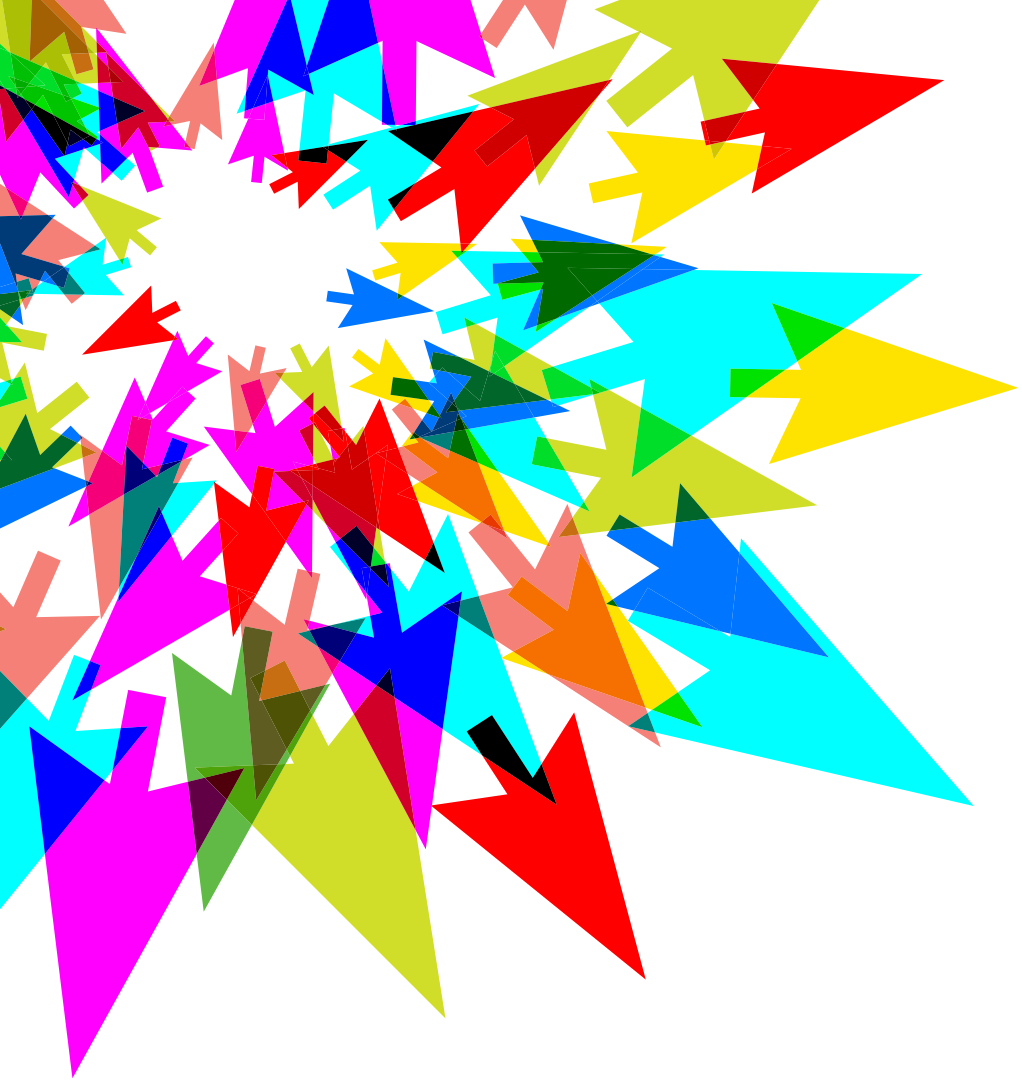


compute canada

A Vision for Research Excellence in Canada

Compute Canada's Submission to the Digital Research Infrastructure Strategy Consultations

Contents

computecanada

A Vision for Research Excellence in Canada

Advanced research computing underpins our national prosperity and is a major ingredient in the recipe that will transform our nation from a resource-based to a knowledge economy. The competitive edge required to develop the workforce for sectors such as life sciences, advanced materials, aerospace, automotive and energy relies on the availability of a robust digital research infrastructure strategy in Canada. Much of the recipe to excel exists today.

Compute Canada is the national organization that plans and oversees Canada's advanced research computing resources, including big data analysis, visualization, data storage, software, portals and platforms for research computing at the majority of Canadian academic and research institutes. It is a national federation of advanced research computing service providers, funded by the federal government, most provincial governments and 34 of Canada's most research-intensive universities and research hospitals.

Together, with our regional partners, ACENET, Calcul Quebec, Compute Ontario and WestGrid, Compute Canada delivers the majority of the large-scale research computing capacity in Canada. In addition, we help to accelerate results and bridge research to industry careers for more than 9,000 researchers across the country. For example, Compute Canada delivered more than 54,000 hours of training in the last two years, thus equipping researchers with the skills required to lead globally in all disciplines and economic sectors.

200 experts accelerating results for more than **8,500** researchers including close to **3,000** faculty members

More than **3,700** peer-reviewed publications, **40** patents, **23** inventions, and **7** companies*

*since 2012

More than **30%** of NSERC's Canada Research Chairs rely on our services

Compute Canada stands out as a model federation as many other economic unions and countries begin to seek ways to share and fund expensive digital infrastructure. We represent the collective national expertise of our best research teams, their international and industry partners and a team of more than 200 technical experts across Canada employed at our Canadian universities and research institutes. Furthermore, Compute Canada also supports our nation's major science investments including: TRIUMF, SNOLAB, Canadian Light Source, Genome Canada, ATLAS and CANFAR.

Scientific computing combines the power of supercomputing and big data to create new and exciting innovations through modelling, visualization and advanced analytics. It is indeed the engine for excellence as much as the CANARIE network is the highway. It is for this reason that Bombardier partnered with Compute Canada for the use of advanced research computing to develop the next generation of clean aerospace. Access to Compute Canada experts and the national platform for advanced research computing was the ingredient Bombardier needed to explore and discover new ways to build their products.

All Canadian researchers have access to Compute Canada resources including disciplines such as medicine, digital humanities, astronomy, computer science, economics, criminology, psychology, physics and engineering. The ongoing renewal of Compute Canada's national digital resource platform will further expand services and enhance offerings in the areas of cloud services and modern systems to meet today's requirements. Currently, the components necessary for a world-class national strategy exist in Canada.

Delivering **54,000** hours of training to more than **11,000** researchers

Serving users at more than **70** Canadian universities

Storing and managing over **15** petabytes of active research data

computecanada

Compute Canada, through its collective expertise and direct link to researchers and the research community, recommends the following to achieve a shared vision of research excellence in Canada:

1. Create an digital research infrastructure advisory council, which would improve coordination of investment and interoperability of technology among current DRI providers, and provide advice to the Government of Canada on publicly funded digital research infrastructure technology

2. Improve coordination of investments in federally funded computing infrastructure to enhance regional economic development activities and industry engagement

3. Ensure sustainable, predictable funding for Advanced Research Computing (ARC)

4. Enhance the coordination for major science and research investments requiring ARC services

5. Implement a flexible funding model and flexible mandate to recognize the central and essential role of ARC services

6. Provide funding to develop software and tools to improve interoperability among different DRI services, enabling researchers to more easily span multiple providers when needed

7. Expand Canada's national data infrastructure to include support for preservation and long term access to valuable research data

Overview of Recommendations

1 Create an digital research infrastructure advisory council, which would improve coordination of investment and interoperability of technology among current DRI providers, and provide advice to the Government of Canada on publicly funded digital research infrastructure technology

Digital infrastructure is costly but has the advantage of being shareable. An analogy would be investing in a multi-million dollar diagnostic tool, and then providing several hospitals with shared remote access. This could also connect to other similar diagnostic machines used for related purposes across the country. Similar to diagnostic technology, advanced computational and data management infrastructure can be costly. Canada therefore benefits from a shared infrastructure model. Improvements to the infrastructure, and to coordination among infrastructure providers, will facilitate outcomes for many activities that require it.

Presently, current organizations within the DRI community adequately represent the user community. And yet, an expert technical advisory council working with the Leadership Council on Digital Infrastructure would bring together publicly funded digital research infrastructure providers, including networking, computing, storage and research data management, greatly improving the coordination of digital infrastructure and resulting technology choices. This advisory council would engage the academic community and beyond, including organizations such as the Canadian Association of Research Libraries, Canadian university campus IT leaders, national research institutes, Shared Service Canada and government and economic development agencies investing in DRI. Through joint planning and recommendations for the technical coordination of infrastructure investments, synergies, gaps and possible duplication of effort would be identified. . Such a Council should thus include all providers and assets in Canada in order to foster coordination and interoperability.

The advisory council would:

- ☑ Map out a comprehensive approach for the technical management of digital research infrastructure and identify areas lacking access, coordination or interoperability;
- ☑ Create a road map for digital research infrastructure to 2020;
- ☑ Develop the technical approaches for using, sharing and managing data (infrastructure) nationally to support the full cycle of data research; and
- ☑ Create a map of resources, searchable by resource type, access mechanism, provider, potential audience, existing usage, and other criteria.

The ability of a researcher to assemble their DRI needs from diverse components and providers is akin to a craftsperson or artist who has an extensive toolbox, and skills to use the tools. DRI tools are available from a diverse group of providers in Canada. Adopting best practices from other models around the globe would help to improve the governance and coordination among infrastructure providers.

Many international models have emerged to address the needed level of cooperation and governance to maximize the impact of investments made in digital research infrastructure.

For example, New Zealand's research infrastructure providers are collaborating to drive capability development in research data at a national level for the benefit of researchers and the economy. Through their 'eResearch 2020' initiative, digital research infrastructure partners initiated the National Research Data Programme (NRDP), a sector-wide initiative focused on lifting New Zealand's capability in data intensive discovery, research data management and digital research methods. Each provider in New Zealand remains an independent organization focused on delivering specific expertise yet connected through a new coordinating framework.

This type of collaboration among providers and funders is the norm across the globe. In the European Union, the DRI ecosystem includes the European Grid Infrastructure (EGI.eu, for federated high throughput/cloud-like resources), PRACE (for high performance computing resources), EuDAT (for research data management), GeANT (for networking) and EUT0 (for uniting key international science platforms). Coordination is achieved through collaboration at many levels from governance to technical operations. Funding from EU, national and subnational sources is leveraged and optimized. Similar ecosystems exist in individual countries across the EU, as well as in the United States, Japan, China and other leading research jurisdictions worldwide.

Increasingly important in every country is that the work of these ecosystem players must be aligned in support of an overarching strategy not just for digital research infrastructure, but for a national science, technology and innovation (ST&I) strategy. Many nations have announced bold strategies that align their digital research infrastructure investments with national aspirations for ST&I. The Government of Canada has recently updated its ST&I strategy, and the digital research community has the opportunity to work with the government to determine how that strategy should inform investments in, and activity by, DRI providers.

With the deep expertise required in networking, computation, software, storage, curation and preservation of data required to host a dynamic DRI ecosystem, global models address the need for multiple technically focused organizations to work together and collectively identify gaps and access challenges.

2 Improve coordination of investments in federally funded computing infrastructure to enhance regional economic development activities and industry engagement

The United States and European Union (EU) have both identified big data analytics and advanced research computing as essential economic enablers. International Data Corporation (IDC) conducted the *Special Study Creating Economic Models Showing Relationship Between Investments in HPC and the Resulting Financial ROI and Innovation and How it Can Impact a Nation's Competitiveness and Innovation Study*. The study reports that one dollar invested in high performance computing (HPC) will result in an economic return of US$356 in less than two years.

Currently in Canada, advanced research computing pursuits are served in a variety of models and funded through a patchwork of programs and mechanisms. These programs and their mandates are currently limited, but in some cases enhance industrial partnerships. Improving and developing new pathways to access the national platform would provide industry with a community of expertise and a variety of technologies not currently available. This would improve our ability to support industry at key competitive stages of development and innovation on a much larger scale.

Small and medium-sized enterprises (SMEs) seek access to advanced research computing capabilities, as well as advice and expertise in this area. Partnerships of this nature are already an important part of the mandates of Compute Canada's counterparts: PRACE and EGI in the European Union and XSEDE in the United States. Canada would be in a better position to serve Canadian research and industrial R&D if the national platform allowed better coordination of access for all types of data intensive research. The combined funding and purchasing power would help scale capacity to more competitive levels and enhance current capabilities. Further, it would allow for improved access to national expertise and research pursuits, sharing of best practices and "one window" access to services, support and partnerships. Academic-led research activities, including industrial partnerships and international collaborations, are largely served by Compute Canada and their regional partners. Funding is provided through the Canada Foundation for Innovation (CFI), which leverages funds from provinces and institutions. The mandate of the CFI is specific and must be academic-led or institution based.

Business-led research is conducted privately or through partnerships with a variety of federal and provincial programs. In some cases, industry obtains access through programs such as the business-led Networks of Centres of Excellence of Canada or NSERC's Collaborative Research Grants.

For some industrial R&D pursuits however, access is limited. While not all private sector advanced computing needs should be served by publicly funded infrastructure, in Canada there are examples of research-intensive pursuits with commercial potential being enabled by early access to expertise and systems. Fusion Genomics, a Canadian research based commercial start-up, was able to develop their diagnostic tool for the detection of early childhood cancers by having access to Compute Canada's advanced computing resources. The resources and expertise made available allowed them to create a testing environment that would not have otherwise been possible. Clearly, facilitating access to computational power and expertise at key moments in the development pipeline can accelerate discovery and help foster growth in the commercialization of research.

3 Ensure sustainable, predictable funding for Advanced Research Computing (ARC)

ARC's infrastructure lifespan is limited; any given system generally requires replenishment or replacement after approximately five years. What is purchased today is influenced by anticipated technology developments and the knowledge of what funds will be available during and after the life cycle of these systems. Compute Canada's current planned concentration of these investments reflects the economies of scale that characterize ARC technologies. Concentrating computational investments into fewer, larger systems provides a number of benefits to Canada and to the researchers and innovators who use the systems:

- Greater value for tax dollars during procurement;
- Improved efficiency of operation (with fewer systems requiring teams of system administrators);
- Improved efficiency of use (more consistent user environments available to larger groups of users, larger systems able to accommodate larger jobs being run by users);
- Coordination among personnel and platforms to investigate and adopt new technologies and new modalities of utilization for ARC;
- Greater technical capacity (diversity of technical resources available on a single system, thereby reducing the need to shift jobs, and increasingly large data files, from system to system; the ability to devote all of a larger system to serve the needs of the most demanding users when required); and
- Increased infrastructure support and services, including broad national-level multilingual support, spanning all of Canada's time zones. Infrastructure support also includes direct access to expert assistance, regardless of which institution the expert is based at.

Given the central and essential role advanced research computing plays in all sectors and disciplines, limiting capacity growth and renewal hinders excellence in Canadian research. Combining purchasing power and enhancing paths to access Canada's national advanced research computing services for extreme computing, big data analysis, robust data storage, software, cloud, platform and portal services maximizes the value of today's investments. This will help to eliminate duplication and provide a path for unused resources to be diverted to other projects while providing priority access as necessary. Attraction and retention of the world's leading talent is heavily influenced by the ability to have access to a robust suite of digital research systems and services.

To be maximally effective, Canada's investments in national-level ARC resources need to be consistent, predictable and ongoing. This will allow for systems to be upgraded or replaced in a timely manner, and for staged deployment of new systems (and retirement of older systems). Ongoing investments also allow researchers to benefit from new system architectures, larger capacity and greater speed, as technologies continue to improve over time.

4 Enhance the coordination for major science and research investments requiring ARC services

As we approach our 150th anniversary, Canada is well-placed to achieve its vision for scientific excellence. Compute Canada is delivering the bulk of services required to achieve research excellence and to support industrial engagement through the provision of traditional compute and innovative cloud services, visualization and domain specific support, software and storage needs for the majority of Canada's research community.

Ubiquitous networking and computing tools, along with expert training and advice on how to use them, have changed the way research is done. Researchers today are able to draw on a variety of infrastructure components, from numerous providers. Today, users of digital research infrastructure are represented through a variety of entities including the Leadership Council on Digital Infrastructure. The user community is asking for improved coordination of services and funding mechanisms. Compute Canada recommendations would significantly improve the management and coordination for a national strategy to support the full research cycle, including preservation and curation (Appendix A).

Building on the priorities set out in the *Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation 2014*, Compute Canada recommends increased coordination and planning between Compute Canada and those government agencies that fund research. Increasingly, this research requires the DRI systems and services that Compute Canada provides. Compute Canada's investments must align with strategic program investments made by these funding agencies in order to realize and maximize their desired outcomes and to attract and retain the best minds, thus continuing to create a nation that innovates and leads.

Today, Compute Canada provides expert advice in connection with CFI-funded projects that require advanced research computing support. This model could be strengthened and extended to other agencies, such as NSERC/CIHR/SSHRC, Genome Canada, as well as to high impact programs such as the Canada Research Chairs (CRC), Network Centres of Excellence (NCEs), and the Canada First Research Excellence Fund (CFREF). Such coordination has already benefited researchers through the use and evolution of shared infrastructure. Requests for DRI investment (above a certain level) are ineligible under Tri-Agency and Genome Canada rules, as well as ineligible under programs such as the Canada Research Chairs and CFREF. However, recognizing these investments require the use of ARC needs to be coordinated.

5 Implement a flexible funding model and flexible mandate to recognize the central and essential role of ARC services

Currently, Compute Canada is funded through two programs administered by the Canada Foundation for Innovation. CFI has been a leader in the funding and creation of a national shareable advanced research computing platform in Canada. CFI ensures that Compute Canada, and the national advanced research computing platform that it provides, support excellent science, creates benefits for Canada and appropriately engages partners in funding the platform. However, the role of advanced research computing has grown from the necessary tool of a few projects in a few disciplines to an essential resource for the majority of today's research pursuits across all economic sectors and disciplines.

A more flexible funding model would facilitate more effective management of the national platform, and provide better value for money and operational excellence. Such a model would further facilitate scientific excellence and benefits to Canada through the access to services that the national platform provides its users. Aspects of this flexibility include combining operating and capital funds to allow overall cost efficiency and return on investment. This model was tested successfully by the CFI in the last round of operational and capital funding submissions; however, these funds still flow from two funding mechanisms.

Compute Canada is seeking a model that recognizes the services that it provides support all of the CFI's Major Science Initiatives program as well as the bulk of ARC services in Canada. Such a model recognizes the longer term operational commitments and capital planning that is required to provide reliable, high quality service to the research community. Scientific and operational excellence can be assured through rigorous periodic performance reviews, without necessarily requiring Compute Canada to compete for funding with the major science initiatives it underpins in Canada.

6 Provide funding to develop software and tools to improve interoperability among different DRI services, enabling researchers to more easily span multiple providers when needed

The current mix of software and related interoperability-focused infrastructure funding to users and to DRI providers generally is not directed at solving broader problems, nor on spanning among DRI providers. Coordination among DRI providers is not sufficient to create the necessary interfaces among services and tools, in order for users to create workflows to solve research goals. Focused expertise must be brought to bear, with goals beyond single research projects or research group goals.

Current research software development for advanced research computing and data intensive research benefits from those computational scientists and developers experienced in creating production-quality software. A broader system evaluation would improve sustainability of investments. Current funding is for teams or individual researchers who may or may not be working toward a scalable, reusable, sustainable model, and in many cases only have access to short term funding. Future software funding should consider whether existing offerings are available to meet the need, mature enough to offer production quality service and their ability to scale on shared infrastructure. “Sustainable software for the 21st Century” is a program operated by the US NSF that demonstrates one approach to this problem.

computecanada

7 Expand Canada's national data infrastructure to include support for the preservation and long term access to valuable research data

Compute Canada manages both the computation and storage requirements for today's active academic research pursuits. While data is active, the creation of backup repositories is automated with meta-data stored, helping to ease the curation process at the end of the data research cycle. Canada requires a strategy to store these and other data indefinitely. Compared to Compute Canada's planned investments in data storage for active research data, this would require a modest incremental investment to support curation of valuable data sets and to preserve them across time, including ongoing file validation and format conversion when needed.

The Canadian Association of Research Librarians (CARL) provides the expertise for the specification of policy and mechanisms for research data retention and preservation. They are working with Compute Canada to gain access to large-scale storage resources, and to benefit from Compute Canada's expertise in big data, data transfer, and other areas. However, it is not within Compute Canada's current mandate to provide long term archival storage. With modest incremental investment, Compute Canada could provide the underlying storage capabilities and related services that would allow researchers, institutions and other organizations the infrastructure needed for curation of data and the implementation of appropriate data preservation policies.

The Government of Canada has announced its commitment to open data in its Digital 150 Strategy. Compute Canada is very supportive of these initiatives. Compute Canada already stores approximately 15 petabytes (PB) of valuable research data and will expand to more than 60 petabytes with recently announced investments from the CFI Cyberinfrastructure Initiative. *(To learn more about planning by Compute Canada for the recent investments in advanced research computing please refer to Appendix B.)*

Specifically, as part of this capital investment, Compute Canada is planning to enhance its data infrastructure with increased capability, accessibility and management tools in order to help researchers exploit the promise of data intensive research.

Compute Canada is committed to continuing to provide future storage capacity, subject to funding. This will enable actively used research data, both old and new, to be stored and made available to the research community. In addition, Compute Canada will continue to accommodate data storage equipment funded and potentially operated or hosted by other partners (e.g., university research libraries, government research agencies). Compute Canada is establishing an open technical environment that will accommodate different patterns of participation.

In particular, through efforts to support the "Challenge 1" Cyberinfrastructure investments of the CFI, as well as supporting the broader needs of the science community, Compute Canada is investing a portion of recently announced CFI funding to build on a pilot implementation of a Research Data Management (RDM) service. This service has the intent of providing data ingest, preservation, meta-data search, data publication and sharing services in a way that can scale to the needs of the broader Canadian science community. Compute Canada will continue to work with CARL and Research Data Canada (RDC) to ensure that these services can enable adherence to best practices by the research community.

Vision for Year 2020

With the implementation of these recommendations, the components will be in place for an integrated approach to data and computational infrastructure of benefit to all sectors of society. A comprehensive system that supports the full data research cycle will allow researchers and their industrial and international partners to compete at a global scale. Through an appropriate coordination framework, each DRI resource provider can apply its deep expertise to maximize all components of the national DRI ecosystem. With enhanced coordination of major science investments that rely on advanced research computing resources, we can ensure that these major investments deliver on their promise. When DRI resource providers are given the mandate and funding to foster interoperability among their resources, providers will create workflows that easily span different resources: from the lab or campus, to national computational resources, analytical facilities, publication archives, and with collaborators. The funding for various components will be organized in an improved model that removes duplication and maximizes investments. Funding and oversight of the various components will be aligned and managed by organizations with the appropriate expertise.

Conclusion

Through modeling, simulation and high performance data analysis, Compute Canada services and platforms are fuelling Canadian companies, government agencies and research investments in Canada. This is encouraging new ways to conduct research and business, develop transformative products, offer services and interact in ways that improve everything from health and safety to productivity and entertainment.

The Government of Canada's recent activities and current examination of its digital research infrastructure is timely and of critical importance to Canada's ability to remain an innovative nation. Compute Canada supports the development of a national digital research infrastructure strategy and looks forward to working closely with its partners to support this exercise. Maintaining and growing current investment levels to keep pace with the science and innovation needs in Canada as well as enhancing coordination and approaches will set the stage for increased competitiveness and research excellence in Canada.

APPENDICES

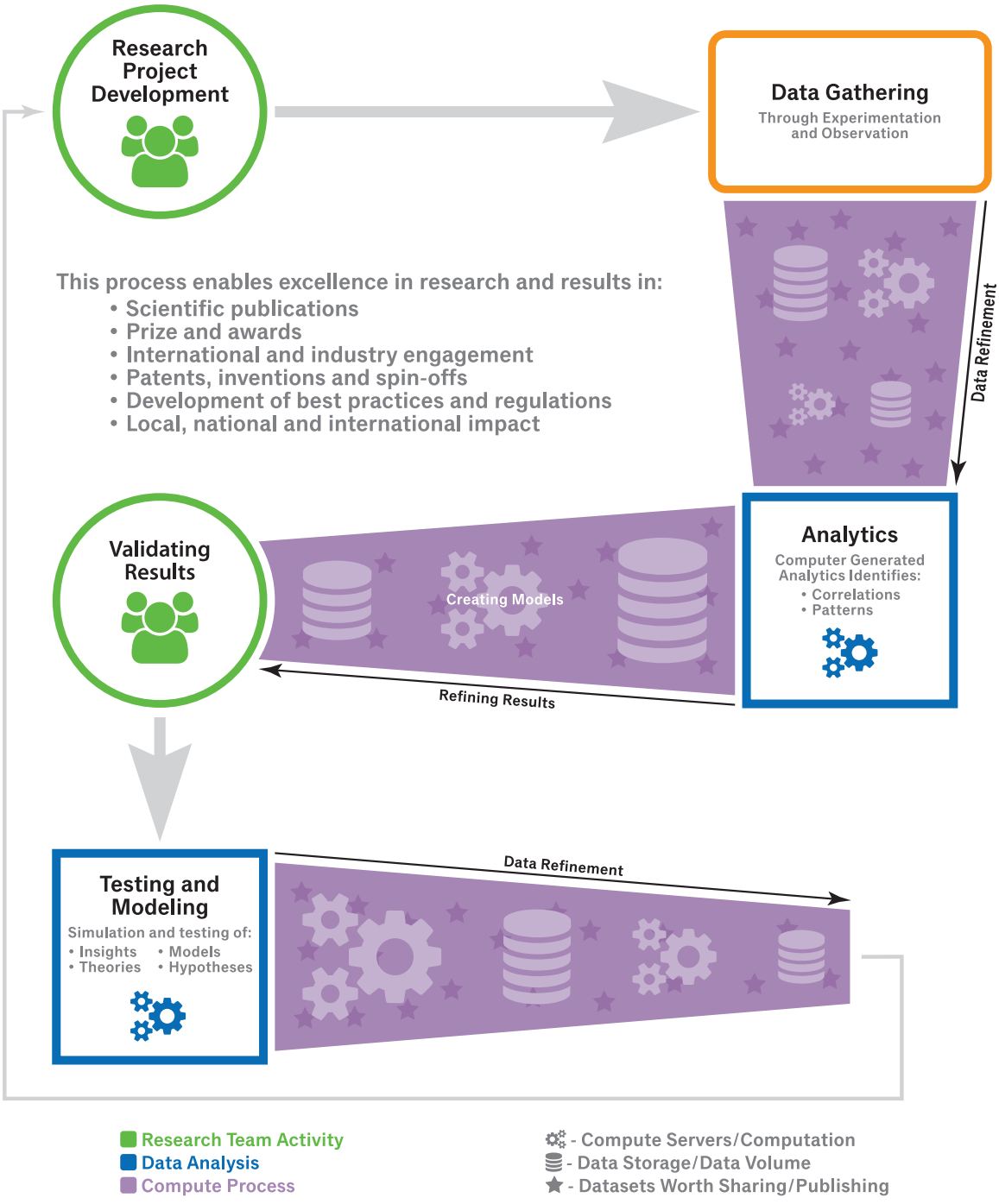
computecanada

Appendix A

Digital Research Cycle

The evolution of data and compute intensive research integrates the classical paradigms of theory and experiment, with the new paradigms of simulation and data analysis.

(see Jim Gray, The Fourth Paradigm)



Appendix B

Current Status of Canada's Advanced Research Computing Platform

Compute Canada provides the advanced research computing resources including big data analysis, visualization, data storage, software, portals and platforms for research computing at a large majority of academic and research institutes. Together with regional partners ACENET, Calcul Quebec, Compute Ontario and WestGrid, we support the *Digital Research Cycle*, in an effort to accelerate the pace of research and allow researchers and innovators to tackle larger problems with greater promise of benefiting our society.

Currently, Compute Canada supports over 2,700 research teams using its systems, comprising over 9,000 researchers and HQP and their international and industrial partners.

Compute Canada is currently leading the broad transformation of its advanced research computing platform, replacing many aging compute and storage systems with modern systems designed to meet today's science needs.

These investments are addressing urgent and pressing needs and are setting the stage for a national platform that can continuously meet the needs of research and innovation in Canada. These new facilities will represent highly concentrated investments, located at four university data centres. The technology refresh program, announced by the government on July 30, 2015, is a key element of a broader, longer-term program of modernization, enhancement, and capacity building. This will be in service to new and existing users, and new models of use, for Canada's Advanced Research Computing infrastructure.

Consolidation of systems

Compute Canada is shifting from operating resources at 27 institutions to half as many over the next few years. Four new systems will become operational as part of the current funding cycle. These new systems will be larger and more capable than the systems they replace, and will provide operational cost savings and economy of scale. Local support — at all of Compute Canada's member institutions — will remain a hallmark of Canada's advanced research computing. Consolidation of systems yields larger and more capable systems, and distributed support yields better capabilities for researchers to do their work.

National-level services, provided by broad cooperating teams. Compute Canada is evolving from regional pockets of expertise and local teams at institutions, towards a model where personnel from across all member institutions work together as national teams. By this mechanism Canada's ARC users will be able to gain access to teams with the appropriate expertise and availability to assist them.

computecanada

Compute Canada's national storage infrastructure is undergoing rapid modernization. Seamless access to data resources will facilitate more complex and robust workflows, while making it easier to share data without unneeded duplication. This new storage infrastructure will have the capacity and features needed for the diverse user base. Features include data isolation for multiple tenants of data resources, object storage access mechanisms, georeplication and backups of datasets, and hierarchical storage management, which will appear to most users as unlimited capacity.

Cloud services for research

Compute Canada is growing its cloud deployment, and integrating cloud services — including storage and job workflows — with larger clusters. Cloud services will host the customized virtual machines needed by some researchers.

Ultimately, Compute Canada envisions utilization by 20182020 that will allow researchers to assemble the system they need from available resources. DRI resource provider cooperation is needed to more easily incorporate resources from diverse providers. Within the advanced research computing environment that Compute Canada will provide, researchers will be able to specify the mix of resources they need — different storage pools and data access mechanisms, long-lived Web or database services, computing capabilities from a single CPU to an entire cluster, visualization and analytic services, and other components. Distinctions that are important today, such as whether a file is local or remote, or the characteristics of a particular high performance computing system, will not be impediments to a researcher to create the system he or she needs.

computecanada

36 York Mills Road, Suite 505, Toronto, Ontario, Canada M2P 2E9

www.computecanada.ca | www.calculcanada.ca | @ComputeCanada