Brief for ISED on the Digital Research Infrastructure Re-organization

Institute of Particle Physics (IPP[1]) and Canadian Institute for Nuclear Physics (CINP[2])

23 April 2019

Innovation, Science and Economic Development Canada (ISED) has announced a restructuring of the digital research infrastructure in Canada. The new infrastructure will consist of a new organization focused on research computing, which will replacing Compute Canada, and the continuation of CANARIE with a revised focus on research networks and cybersecurity. In the interim period, when the new research computing organization is being established, ISED will maintain the existing Compute Canada sites and continue to support CANARIE. The purpose of this briefing note is to provide ISED with input from the Canadian subatomic physics research community that will help guide the organizational and management structure of the new organization. Our goal is to ensure that the new organization effectively provides the critical infrastructure and related support to the entire Canadian research community in a manner that, at minimum, is consistent with researcher expectations and international norms.

Subatomic physics includes the fields of nuclear and particle physics with the common goal of understanding the fundamental nature of the universe. Our community is represented by the Institute of Particle Physics (IPP) and the Canadian Institute for Nuclear Physics (CINP). We are involved in a wide range of international projects, such as the ATLAS project at CERN, as well as in projects within Canada at TRIUMF and SNOLAB. Researchers in our community also use the Perimeter Institute and Arthur B. McDonald Astroparticle Physics Research Institute. Our research is conducted in Canada and abroad and involve both small and large national and international scientific collaborations.

Computing is an integral part of our research and its use ranges from medium to large scale computing; storage of large data samples; and cloud and high performance computing. High-speed networks are also an essential component as they enable us to communicate with our peers, e.g. via video conferencing, and transfer large volumes of data around the world. Historically, we have funded these facilities from a variety of sources: directly from CFI and NSERC, or through allocations of resources from Compute Canada. Today our community uses many thousands of compute cores and many tens of petabytes of data storage. Our computing and network requirements will continue to grow over this decade and beyond. Our goal is to ensure that ISED understands the requirements of our community when it establishes the new research computing organization and can develop an organizational structure that is responsive to the various needs of each research community. The subatomic physics research community is willing to commit time and effort to help ensure the success of the new organization in its effective delivery of its resources to the entire Canadian research community.

---

[1] Institute of Particle Physics (Canada), Director J.M.Roney (www.ipp.ca)

[2] Canadian Institute for Nuclear Physics, Executive Director G. Huber (www.cinp.ca)

This document was prepared by the Joint IPP CINP Computing Task Force: A. Garnsworthy(TRIUMF), D. Gingrich, G. Huber, C. Jillings(SNOLAB), R. Lewis, J.M.Roney, R. Sobie, R. Tafirout(TRIUMF)

The appendix to this document describes a sample of some technical challenges faced by our community in using the current digital research infrastructure system. Our experience with computing resources provided by our international partners and the Canadian ATLAS Tier-1 computing facility indicates that Compute Canada, if properly organized, can do better.

In planning for the new research computing organization, we recommend the following:

1. Conduct extensive consultations and reviews with researchers within the various research communities in order to fully understand the technical requirements of communities and the challenges they have been facing. Such consultations should include international experts with extensive experience on similar systems. To date, consultations have been limited to high-level administrative personnel at the universities who are typically removed from the day-to-day realities of users of the digital research infrastructure.
2. Treat research computing as a national resource with an expected uptime of nearly 100% and involve stakeholders in failure-point analysis during specification of new systems.
3. Ideally have the new organization led by a highly-respected researcher with experience as a user and who has vision and expertise in research computing and can inspire the Canadian research community. Our experience at SNOLAB, Perimeter Institute, and TRIUMF has shown that the laboratories and institutes are successful when led by a highly respected and passionate researcher dedicated to making the facility work.
4. Create a Research Advisory Panel that effectively engages with the new organization's management on an ongoing basis and which periodically reports back to ISED. One effective model for this is the regular reviews provided to TRIUMF from its Advisory Committee on TRIUMF (ACOT).
5. Organize computing infrastructure systems configurations, architecture and support team resources according to the functional requirements of an explicitly identified set of research communities. The current Resource Allocation Committee (RAC) structure could serve as a mechanism for identifying those different research communities.

   Subatomic physics, for example, makes extensive use of high numbers of compute nodes and some projects require 24/7 access to specifically configured computing resources as well as extensive network resources to fulfill computing obligations in international collaborations. Other fields require higher I/O or systems configured for highly parallelized code, and yet others have higher storage demands.

   This recommendation will result in systems that are more manageable, both technically and administratively, and will be more cost-effective. Configuring systems designed for field-dependent performance instead of general-purpose facilities, and engaging teams with specialized expertise, will better meet the needs of all research communities.

   The support teams, who will be knowledgeable about the specific requirements of each research community, will provide critical interface between the hardware and related systems software. Providing researchers with such teams will ensure the most effective use of the resources by those communities.
6. Introduce data archiving as part of the mission of the new organization. It is important in many fields to ensure the long-term storage of unique data. Compute Canada did not explicitly have the mandate to provide long-term data archiving and we see no other solution for this need, which spans many research communities, other than to have it stated explicitly as part of the new organization's mandate.

In implementing these recommendations, and in order to avoid the challenges observed with the current digital research infrastructure, the new research computing organization will:

- Adopt best practices when managing its resources and interacting with its user community. Establish lines of responsibilities between the site personnel, the domain-specific teams and the management with a working and transparent ticketing system that interacts effectively with the users. Ensure redundancy in expertise and skill-sets by cross-training existing staff to avoid any 'single-point-of-failure' associated with planned or unplanned leaves. Use a coherent set of rules and robust procedures across all sites used/allocated/assigned to by the particular research community generating the tickets.
- Provide the key performance indicators for each site that are reviewed by the Research Advisory Panel and made publicly accessible. For example, a quarterly report provided by the new organization should include the CPU utilization of a site, a list and reason for the downtime of each site, and percentage of resource allocation actually used by users compared to those allocated.
- Establish an enhanced resource allocation process that ensures the RAC members are knowledgeable about the type and availability of resources. Currently resources are often allocated on incompatible systems or systems that were heavily oversubscribed because of a lack of information provided to Resource Allocation Committee members. The RAC should also take into account hardware decommissioning schedules, which is particularly important to maximize the stability through the lifespan of projects with a long life-cycle. In principle, if Recommendation 5 is adopted, this would be straightforward.
- Ensure systems are as accessible as possible to international collaborators of Canadian researchers within the allocations granted a given project.
- Work closely and collaboratively with CFI and NSERC, as all are critical for the success of any given project.

The new research computing organization, with an enthusiastic and effective CEO, has an opportunity to make an impact in new and innovative areas. The new organization could explore new initiatives with industry, for example:

- Enabling the use of commercial cloud computing resources by providing funds and resources/teams that could, for example, meet peak demands that exceed existing resources. The commercial cloud systems are extremely reliable and their technical teams are highly responsive to their customers' requirements. Their systems have been demonstrated to be equivalent to the existing research computing systems in Canada. The new organization and researchers in Canada would benefit by strengthening the ties with the commercial cloud computing industry;
- Invest in the development of research software and data management. Our preference is to provide funding to researchers through a competitive peer review process that draws on the established experience of NSERC. In fact, it would be most efficient to have such competitions managed by NSERC. Funds in these areas should be focused on our highest priority projects that have significant international impact based on peer review.

The Canadian Subatomic Physics community is internationally recognized as having a leading role in providing computing resources, evaluating emerging technologies, such as cloud computing, and pushing the limits of our research networks. We believe the new research computing organization is an opportunity to establish Canada as a leader in these areas and we are willing to commit our efforts to ensure its success.

Appendix

For reference, we provide examples below of some technical challenges faced by our community when using the current digital research infrastructure in Canada, all of which would be addressed by adopting the recommendations in this brief.

- A number of projects have reported loss of critical/unique data. The current backup systems fail to guarantee the security of the data. The new organization needs to work with the users to create near-fail proof systems for data integrity.
- Sites are frequently offline, often for extended periods, and often unscheduled and with minimal information sent to users. Each site needs to be designed with full redundancy. For example, computing systems at the CERN Laboratory in Geneva or the commercial compute clouds are never offline to the users.
- Users are reporting that logging into the interactive systems takes an unreasonable length of time (often up to 10 minutes). The new organization needs to build responsive and performant systems for the users.
- RAC allocation has not always taken into account the long lifespan of typical large projects requiring stable access to resources for several years, and planned decommissioning of systems. It is critical to provide stability over project lifespan.
- Users are finding that their queries for assistance and error reporting are not followed up. There is a general impression that the management and the technical teams do not effectively communicate with each other on user issues. Further, it often seems that the technical teams manage themselves without direction from management.
- The existing sites seem to be operating with a small number of overworked experts. This often results in crisis management - solving the latest emergency. It may be necessary to reallocate some the existing staff to make the current systems more reliable and redundant.
- Many users report that the resources are unable to meet the RAC allocations - either there are not enough resources or the award was given on the wrong site. This may be remedied by the interim funding from ISED.
- It often occurs that critical system managers and other experts do not have back-up when they go on holiday or other leave, leading to ‘single-point-of-failure’ problems that significantly impact the productivity of the research communities. This problem can be avoided by cross-training existing staff.