Digital Research Infrastructure for Canadian Astronomy

Catherine Lovekin (Mount Allison University), Pauline Barmby (University of Western Ontario), J.J. Kavelaars (HIA), Adrian Liu (McGill University), Erik Rosolowsky (University of Alberta), Kristine Spekkens (RMC), James Wadsley (McMaster University)

Astronomers have long been a major user of DRI infrastructure in Canada. From 2017 to 2019, astronomers submitted 5% of resource allocation requests to Compute Canada, but used 10% of the processing capacity and 20% of storage (Compute Canada Statistics, courtesy E. Fuenmayor). However, this heavy usage has been essential for the successes of Canadians in the field. Internationally, Canadian astronomers are recognized for high-impact research as measured by mean citation counts (Crabtree, 2019). In fact, Canadian astronomy and astrophysics was listed as in the top 1% in the world; the only Canadian STEM field to be so highly ranked (Council of Canadian Academies, 2018). Much of this research depends on DRI, including numerical simulations run on Advanced Research Computing facilities provided through Compute Canada, Research Software authored and supported by our community, and over 40 years of leadership in Research Data Management by the Canadian Astronomical Data Centre (CADC). Canadian astronomers are world leaders in developing analysis tools for observational data and large-scale simulations. In this white paper, we address the current and future needs of the Canadian astronomical community, which must be supported by NDRIO for astronomy to maintain its current level of excellence. In parallel to this whitepaper, we have also submitted the *Canadian Astronomical Society's 2020 Long Range Plan for Canadian Astronomy* covering the next decade, which provides the broader context for these DRI priorities.

Current Issues:

Answers to fundamental questions in astronomy and cosmology require not just traditional observations or calculations, but also need enormous data sets and large-scale computational simulations. Many theoretical programs require formidable computing resources, both to perform simulations or calculations and then to store and analyze the output. Several of the future observational facilities and programs recommended the Long Range Plan for Canadian Astronomy will produce exabyte-scale data products, components of which need to be distributed electronically to the community and also archived in perpetuity. Furthermore, fundamental analysis tasks such as stellar classification, redshift determination, and cross-correlating multi-wavelength sky surveys have become so complex that they now rely heavily on approaches involving astroinformatics, machine learning, and artificial intelligence. The computational facilities available to Canadian astronomers process petabytes of data per year from current experiments. CADC staff provide a leadership role in the development of data standards within the International Virtual Observatory Alliance, the international body that provides a uniform research data management environment for astronomical data. They are also bringing astronomy research computing into a cloud-based environment that crosses boundaries across wavelength and research domains. Access to DRI is fundamental to the ability to carry out modern astronomical research, but the field faces several DRI-related challenges in the coming decade.

Challenge 1: Advanced research computing is under-resourced.

Internationally, Canada does not rank well on the list of countries hosting the most powerful supercomputers. Niagara is currently ranked at 82 (#69 in 2019 and #53 in 2018). Our access to HPC resources is approximately a factor of 5 lower than flagship systems in other countries. Astronomers must apply for large blocks of time through the current Compute Canada RAC program, but the size of allocation required may not be sustainable as demand grows. For example, 3D simulations of stellar convection with nuclear burning required 3700 core years (led by Falk Herwig, UVic), or about 6% of Niagara for a year. The largest cosmological galaxy simulations require even more core years, with the Illustris-TNG simulation (run on the German S-MUC system in 2018) requiring 15000 core years. This would be equivalent to ¼ of Niagara.

Currently, some astronomers are forced to find time on systems outside of Canada. For example, the recent Canadian-led "Tian-Nu" simulation (Emberson et al., 2017) ran on over 13,000 nodes on computers in China, highlighting the need for increased capacity in Canada. In order for Canadians to remain competitive in the realm of theory and simulations, NDRIO must begin a program of sustained investment in computational capability. To maximize efficiency, these systems should be as integrated as possible, so that researchers are not faced with technological challenges of transferring calculations from one facility to another.

Challenge 2: The rapid growth of astronomical data from both observations and simulations.

The next generation of observational facilities will be producing >100 PB/year by the end of the decade, which will require computational capacity to process these data and large amounts of data storage. Initiatives such as the Square Kilometer Array (SKA) Regional Centre will require huge amounts of storage and compute capacity (to be addressed in a separate white paper; Spekkens et al. 2020) as part of an international agreement for Canadian participation in this project. In addition, the Canadian Hydrogen Observatory and Radio Transient Detector (CHORD) has been fully funded by CFI, on the condition that the required computational resources can be secured. These resources are expected to be substantial; an estimated 700 TB of storage per year and 1200 cores of compute resources.

The science results of these new initiatives will also join with the existing archives of astronomical data that stretch back to the dawn of digital astronomy in the 1980s. These archives of observational data can produce useful results for decades after the original observations are collected. Archives are also important as so many objects are variable and/or evolving. When observations of an object are collected, there is no guarantee that those data can ever be repeated. Archives are thus important for re-analysis, but also for discovering how objects evolve over timescales of decades. Observational data often requires significant post-processing as well as databases and catalogues so information can be found. Access to the data requires good network connections, and processing of the data requires high-throughput computational facilities. These archives also act as software repositories for both simulations and data-reduction pipelines.

Canadian astronomers are fortunate to have a facility like the CADC, which allows for long term archiving of large data sets. According to the CADC website, in 2019 astronomers in Canada and around the world downloaded over 2 Petabytes of data from the CADC. However, upcoming projects such as the Large-Scale survey of Space and Time (LSST), the SKA and CHORD, are expected to produce many times this volume of data annually.

Challenge 3: The current Compute Canada model for allocation of resources is inefficient for astronomy.
Many HPC initiatives require concentrated resources for short time periods, rather than a uniform share of resources distributed over the entire year. In addition, when introducing a new code or set-up, the computational requirements for testing scaling can be demanding. However, applications to the RAC must describe how a researcher's code scales, which cannot be determined without a successful allocation through the RAC.

At the moment, there is no mechanism to apply for time outside of the normal RAC cycle. A mechanism for fast turnaround applications requiring resources above and beyond the current RAS throughout the year would be beneficial for testing codes and for dealing with challenges or opportunities that arise unexpectedly, as Compute Canada has done with the response to COVID-19.

The timing of the RAC is also challenging given the typical cycle of students and postdoctoral researchers. Applications are submitted in November, and the allocation begins in April. New students and postdocs typically start in September, which means PIs must either be able to plan who they will hire and what their project will be 10 months before their arrival, or wait 8 months for the project to be granted time. Again, many researchers are forced to rely on international collaborators to get time on systems in other countries to fill the gap.

Future State:

From June until September 2019, the CASCA-Computation and Data Committee (CDC) conducted a survey of CASCA membership to project the community's usage of computing and data resources over the next decade. We summarize the results of the survey below with the raw data available here. The results are discussed in more detail in the CDC report of September 2019[1]. Of the survey respondents, about half are current users of Compute Canada resources. Users were overall satisfied with Compute Canada services, although many noted limited access due to long queue times or cluster instability.

We asked users to anticipate the growth of their storage and processing requirements over the next 5 years. Nearly all respondents (90%) indicated a three- to ten-fold increase over their current usage in the next five years. Given their projected requirements, we estimate that the astronomical community will need 100 Pflop years of GPU time, 100 Pflop years of CPU time, and 75 PB of online storage by 2025. Note that these numbers do not include resources for specific projects, such as the SKA Regional Centre.

[1] Available online at https://casca.ca/wp-content/uploads/2020/04/CDC_final_Sep30.pdf

We also asked respondents regarding their priorities for new investment that would support community efforts. They were asked to rank the priorities of the following facilities, for which we report the mean rank and standard deviation of ranks from first (1) to last (4) priority:
● Astronomy-focused Computing Centre (2.14, 1.09)
● New Data Intensive Computing Facility (2.26, 0.94)
● New High Performance Computing Facility (2.74, 1.24)
● Next Generation Software (2.82, 1.01)

Of note, the responses for a New High Performance Computing Facility were bimodal leading to the high standard deviation. Respondents tended to rank this priority either first or last.

NDRIO has a mandate to serve the Canadian research community, but this community does not function in isolation. It is important to recognize the impact of international collaborations, and NDRIO should work to make connections with international counterparts.

In addition, Canadian astronomy has often made in-kind contributions to telescope projects, offering data storage and support to get data access and telescope time. This kind of collaboration is critical to the success of the astronomical community, and there need to be mechanisms in place to support these kinds of international links.

As discussed in the current state, Canada currently lags in computing infrastructure compared to many nations. We need investment in large scale compute facilities to remain competitive, and these facilities need to be continually renewed.

In addition to facilities, NDRIO needs to invest in support staff. These staff members are essential to running the facilities, but can also contribute to the research projects. CANARIE offers programs to bring software developers into research groups, and NDRIO should offer similar programs.

How to Bridge the Gap:

Current funding levels need to be increased. In order to keep up with demand, the available computational resources need to grow exponentially. If we want to maintain our current reputation internationally, our resources need to grow for us to keep up.

A key component of the DRI ecosystem are the support staff. These staff members can train researchers, and ideally help with making software more efficient, which will allow researchers to use the available resources more efficiently. Currently, many of the CC staff are supported by provincial matching funds on 1 year contracts. Long term stability will help recruit and retain HQP.

NDRIO must make it clear to researchers what additional services will be offered, or it will just be seen as Compute Canada with a new name. Consistent and ongoing renewal of hardware, enhanced support services, more diverse ways to access resources, and enhanced data storage capabilities are all ways NDRIO can stand out.

Canadian astronomers are world-leading researchers, in a large part because of the computational and data resources available. It is vital that NDRIO continues to support these endeavours with increased computational and storage resources.

References:

Council of Canadian Academies, 2018. Competing in a Global Innovation Economy: The Current State of R&D in Canada. Ottawa (ON): Expert Panel on the State of Science and Technology and Industrial Research and Development in Canada, Council of Canadian Academies.

Crabtree, D. 2019. Canadian Long Range Plan for Astronomy and Astrophysics White Papers, LRP2020. Online at https://www.zenodo.org/communities/lrp2020, id.14

Emberson, J., Yu H.-R., Inman D., Zhang T.-J., Pen U.-L., Harnois-Déraps J., Yuan S., et al., 2017, RAA, 17, 085. doi:10.1088/1674-4527/17/8/85