

## **Computation and Data Committee Report to the CASCA Board, May 2015**

Current Committee membership:

James Wadsley (McMaster) (Chair)	Term ends: 31 June 2016
Hugo Martel (Laval)	Term ends: 31 June 2017
J. J. Kavelaars (HIA/NRC/CADC):	Term ends: 31 June 2017
Erik Rosolowsky (Alberta)	Term ends: 31 June 2015

Erik Rosolowsky is willing to serve for an additional term (to June 2018).

### **Computing in Canada**

#### **Current Status of Compute Canada**

Compute Canada (CC) now has a full management team led by CEO, Mark Dietrich. The CSO (Chief Scientific Officer) is Dugan O'Neil (Physics, Simon Fraser) and there is a recently appointed CTO (Chief Technical Officer) Greg Newby (Academic IT and supercomputing background). The CC Advisory Council on Research (ACOR) includes James Wadsley (McMaster), of this committee and Robert Thacker (SMU) as chair. New-look website: <http://computecanada.ca>

#### **New CFI Funding**

CFI announced new funding with 4 installments. Challenge 1 is research data proposals, focusing on big data-style research platforms (including personnel but not hardware) and Challenge 2 is hardware. The applications and decisions are staggered in time covering a total period of roughly two years.

The first is a \$40M “urgent” refresh of Compute Canada's aging equipment (Challenge 2, stage 1). The dollar amounts include 40% CFI, 40% province and 20% other (often vendor discounts). The proposal will result in one new large parallel facility (comparable to GPC) with of order 32,000 cores in Toronto and 3 smaller machines with 10-16,000 cores each and some GPU capability located in Waterloo, Burnaby, B.C. (Simon Fraser) and Victoria. This plan is nearing final approval (June 2015) but installation will not occur until 2016, about 10 years after the last major funding of HPC by CFI. The current Compute Canada plan is to maintain some existing systems at least until all the new systems are on-line. The remaining larger older systems will be at 5+ years old at that point and too expensive to maintain for an extended period. There is real possibility that the refresh will ultimately entail a net decrease in processing capability. In comparison, our competitor nations tend to increase net capability by factors of 4 or so every 3 years. Thus compared to our recent peak in capability, around 2010, in 2019 Canada's relative capability will have dropped by about an order of magnitude without further investment. There is considerable effort within Compute Canada to push for a regular refresh process that would keep Canada more competitive. However, it is not clear what lobbying role Compute Canada should play. CFI, in particular, tends to see Compute Canada as a manager of infrastructure rather than a representative of the research community. ACOR plays a role here but as an advisory group it tends to be responsive rather than pro-active.

There will be a second stage to Challenge 2 in April 2016 for which the total expenditure is unclear but may be comparable to stage 1. It is designed to meet research data needs associated with Challenge 1 described below. CFI's commitment to storage for research data needs is still unclear. In particular, many disciplines, including astronomy, have growing data needs that are outstripping the storage and processing capability of institutions. Compute Canada's current storage capability is just a few PB and the stage 1 will upgrade that to ~ 10 PB. It follows the CFI decision on the research data proposals.

Regardless of the selected Research Data projects, the astronomy community should pay close attention to this process and make sure our storage and data analysis needs will be adequately served. In particular, many observational astronomers are making use of the Compute Canada OpenStack cloud systems. The growth in cloud systems for processing observational data provides a good national platform for dealing with the large data challenges that future instruments would impose. Given the outcome of Challenge 1 calls, developing more data-intensive computing infrastructure would be a clear need within Challenge 2.

Challenge 1 focused on research data infrastructure projects. These are discipline specific and there are two major astronomy-related proposals being submitted. The deadlines for NOI are May 22<sup>nd</sup>, 2015 (1<sup>st</sup> competition) then April 2016 (2<sup>nd</sup> competition). If invited, full proposals are due on October 16<sup>th</sup>, 2015 and October 2016 respectively. These competitions encourage the development of general software platforms for research data including storage, processing, analysis and distribution. The total funding is \$10M (CFI) per competition with any one proposal to be given no more than \$2M from CFI with matching from provinces and in-kind as usual. The primary targets for the money are personnel (programmers, software developers, analysts) and software expenses. Hardware is not ordinarily eligible but there will be a process (Challenge 2, stage 2) designed to make sure Compute Canada can provide the required hardware infrastructure. CFI is seeking to develop general tools and software platforms that can benefit the wider community within each research discipline and also be transferrable to other disciplines.

The first astro-related NOI is centred around CANFAR: migrating to Compute Canada and upgrading that platform. The migration part of the process already has significant financial support from NRC and the cooperation of Compute Canada. The CFI proposal will extend CANFAR use cases to include simulated data, to be able to handle (store, distribute and analyse) larger datasets and novel processing modes including common pipelines for simulated and observed data. PI: Pritchett (U Vic)

The second astro-related NOI has as lead institution the University of Toronto. It will develop pipelines that focus on climate and planets, large scale structure and big data, gravitation and rapid signal processing. PI: Gaensler (Toronto)

We recommend that CASCA be engaged in both these proposals and encourage complementarity between them.

## **CANARIE**

Much of the development of the CANFAR platform has been funded through CANARIE and the current NEP support is winding down. In the 2015 budget, the Federal Government renewed the mandate for CANARIE and provided five years of budgetary support. There is a strong possibility to renew support for CANFAR via an imminent call for proposals. CANARIE generally seeks to fund network related tools and capabilities, not science specific ones.

There is substantial overlap between the initiatives that CANARIE supports and the CFI Cyberinfrastructure discussed above. The interpretation of the CDC has been that CFI is more interested in domain specific applications though both streams represent possibilities for continued CANFAR funding.

Since CANFAR infrastructure is the foundation of CADC archival holdings, there is a clear need for

funding through one of these streams. Without support from either CFI or CANARIE, the archival legacy of our telescopes is at risk. It is becoming necessarily more difficult to operate an open science data archive in the IT environment of Government of Canada servers. With increased security measures in place, it remains important to host CADC holdings and processing entirely within research infrastructure like Compute Canada.

CANARIE also has supported the CyberSKA portal, which has become an avenue for Canadian work on the SKA Science Data Processor. CANARIE funding thus represents a good channel to maintain Canadian contributions for the pre-construction phase of the SKA.