

## **Computation and Data Committee Report to the CASCA Board, June 2014**

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 2014
Erik Rosolowsky (Alberta)	Term ends: 31 June 2015

We welcome Hugo Martel to the CDC committee. Other current members have indicated that they remain willing to serve.

### **Computing in Canada**

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

Compute Canada (CC) has now appointed a permanent CEO – Mark Dietrich, formerly the deputy-chair of the CC board. He officially started in April, 2014. Mark was formerly CEO of the Ontario Society of Professional Engineers and prior to that worked with the Ontario Research and Development Challenge Fund (ORCDF) and was involved with supporting Ontario's HPC consortia. Though Mark has no research experience per se, he is felt to be more engaged and consultative than the former, interim CEO. Jonathan Dursi is no longer the CC CTO (Chief Technical Officer) and this position remains unfilled. The CSO (Chief Scientific Officer) is Dugan O'Neil (Physics, Simon Fraser). The CC Advisory Council on Research (ACOR) includes James Wadsley (McMaster), of this committee and Robert Thacker (SMU) as chair. Website: <http://compute canada.ca>

#### **New CFI Funding**

As reported by this committee, new investment in HPC and other computing support has been a problem for years. A welcome development is a new Cyberinfrastructure Program from CFI: <http://www.innovation.ca/en/OurFunds/CFIFunds/CFIcyberinfrastructureinitiative>

The CDC highly recommends that the board read the short summary page linked to above. It might also be distributed to the CASCA membership.

This brand new initiative (first appearing on the CFI website in June 2014 and in documents circulated to University Research Offices in May 2014) resulted from in-house consultation by CFI with selected researchers rather than lobbying by CC or others (e.g. Jan 2014):

<http://www.innovation.ca/en/Media/News/CanadaFoundationInnovationworkshopcyberinfrastructure>) Researchers hammered home the need to replace and upgrade existing capacity and to support the surge in big data and CFI finally paid attention. Note that this is a real shift. CFI said only last year that researchers had previously told them everything was fine for years to come. This indicates that CFI's consultation process is flawed and also that CC and other groups are not lobbying effectively to supplement CFI's internal processes.

The title cyberinfrastructure is a deliberately broad term selected by CFI to allow for non-traditional facilities (e.g. cloud) and software development. The initiative has two components – a renewal and upgrade of CC HPC facilities (essentially all major sources of compute cycles in Canada) and a call for proposals for data-projects – both equipment (mostly storage) and support (e.g. software development and staff). The data-related proposals will be selected via a competition (NOI Jan 2015). Data-project equipment is expected to be procured by CC and housed in CC managed datacentres for the most part. However, the data-projects can also ask for overhead/people money. CC itself is expect to use existing

operating funds to pay staff/manage the equipment. CC is asked to submit its renewal plan for April 2015 and CFI will interleave the two parts and look for consolidation. In particular, CFI seems to assume data analysis will use CC systems.

The total budget is \$50M (towards 40% of total project costs, thus requiring matches from provinces and other partners). CFI does not have a preset division between renewal of HPC and data-related applications and the split will need to be argued to some degree. There will also need to be some rationalization of the proposed data-project proposals and the renewal of CC HPC. For example, CC systems may be enlarged/modified to accommodate needs for data processing. This division between CC HPC renewal and data seems a bit artificial – simulations also produce big data. There is also a risk that the data needs of less organized or smaller groups may be overlooked. Therefore, we would hope for considerable storage capacity in addition to that earmarked for major projects.

CFI is undertaking consultations on this process with researchers and will visit many Canadian institutions/cities (including: Montreal, Toronto, McMaster, Waterloo, Western, Vancouver, Calgary, Edmonton, Winnipeg) between June 5<sup>th</sup> and June 12<sup>th</sup>. Unfortunately, much of this overlaps with the CASCA meeting. Hopefully, astronomers not at CASCA will make an effort to attend these sessions. It is expected that CFI will use these consultations to refine/tweak the plan and timelines. The plan will formally get underway in the fall with calls for proposals/NOI (October 2014).

Compute Canada is starting to move and ACOR will help put together a community consultation process. The tentative name for the CC renewal proposal is the Sustainable Plan for Advanced Research Computing (SPARC). From the CDC's perspective, it is very important that this be grass roots from the researcher base upwards, via town-halls at institutions, existing consortia and regional fora. Ideally an expert researcher panel representing both regions and disciplines would draft the renewal plan.

As part of the process, CC will solicit whitepapers from disciplinary organization (e.g. Astronomy/CASCA). The renewal proposal must balance attempts at novelty (something CFI has asked for in the past) with cost-effective provision of accessible resources. For example, GPU (graphics cards) are powerful but most users are not positioned to make use of them for everyday computing. Similarly, Cloud Computing is a popular idea, but if we use 3<sup>rd</sup> party HPC (e.g. Cloud from Amazon) it cuts-off our HPC as soon as the nominal funding period ends and doesn't leverage our in-house staff expertise. 3<sup>rd</sup> party cycles are not cheap either. The renewal proposal must be sensitive to provincial concerns and make good use of existing staff and infrastructure (e.g. machine rooms). CC has tended to poll researchers rather than engage with them and putting SPARC together will be a real test of CC's ability to work with the community rather than around it. The previous CC strategic plan exercise was not encouraging in this respect.

### **Astronomy: Impacts and Action Items**

There are several Astronomy affiliated groups/projects with substantial big data needs, e.g. CHIME, CANFAR, CyberSKA and the CADC. In particular, both CANFAR and CyberSKA intend to put in CFI data project proposals at this stage. The CDC proposes that CASCA be highly supportive of these initiatives. A key consideration is whether things could be usefully combined. For example, should there be a single data-project initiative spearheaded by CADC that Canadian Astronomy supports? Data needs can require local file systems due to the very high data rates in many cases. In this case, data handling might be staged with local processing and then a longer term repository accessible to the community with associated HPC resources to do analysis. There should be a census of these needs

and then coordination to make sure that centralized data storage/analysis and project specific components can meet the combined need.

On the simulation side, Canada has been falling behind in many respects. As international systems have grown larger, competing groups have run larger simulations. The available cycles per capita around the world doubles almost every year. Consequently, 1000's of CPU years are now being applied to single simulations on 10,000+ cores. Typical simulations run on Canadian systems have not kept pace. Partly, this reflects a shift associated with the lack of renewal of Canadian systems. Canadian users have tended to work on non-Canadian systems and/or self-limit to more modest simulations. The number of Canadians working at the high end in most disciplines (i.e. outside Astronomy as well) is fairly small. These factors help explain the apparent contradiction whereby Canada can have relatively low computing per capita but many users can still be content with the available cycles – even the handful of users doing very large simulations – and report to CFI that their needs are met. This points to an ongoing slide in Canada's HPC competitiveness. Internationally, many fields are moving more heavily into high-end HPC and doing qualitatively new science while Canadian researchers are not. A lack of support for programmers and data-management has been a factor limiting Canadian researchers and their willingness to move to the cutting edge. For example, simulators, like observers, do not have the capability right now to handle the data volumes that cutting edge research produces. CFI tends to be backward looking – supporting only established excellence. It does not invest enough in developing future excellence. CFI has expressed a willingness to invest more in software and other infrastructure that supports HPC and this may be critical to reviving Canada's competitiveness.

CASCA should contribute a white paper as part of the SPARC/CC renewal process. This will require polling the community to find out their needs over the next 5-10 years. It must also advocate for a general pool of storage and compute power to allow for unplanned needs. It should also support the development of new capacity with respect to simulations and analysis (e.g. methods and software) so that Canada can compete in future.

CASCA and its members should participate in appropriate fora such as ACOR and town-hall events organized by both CFI and CC. Fortunately, ACOR is well covered at the moment with two astronomer members.