



**White paper presented by the  
Canadian Society for Digital  
Humanities / Société canadienne  
des humanités numériques  
(CSDH/SCHN) in response to  
Compute / Calcul Canada's  
2016 Call for White papers for  
Sustainable Planning for  
Advanced Research Computing  
Phase II (SPARC2)**

As part of the SPARC2 process, we welcome the opportunity to contribute a white paper that reflects on Compute Canada's ongoing support for humanities research since the Canadian Society for Digital Humanities / Société canadienne des humanités numériques (CSDH/SCHN) presented a white paper to Compute Canada in 2014 in response to the first SPARC call. In the intervening years, we have been very glad to work with Compute Canada to help raise humanities and social science researchers' awareness of Compute Canada's advanced research computing infrastructure and support. We look forward to continuing to work together in this vein.

This white paper, like the first, results from a call for input from the community at large. The original paper outlines the infrastructure challenges faced by digital humanities scholars working in Canada. This second white paper reflects on changes since 2014 with a particular focus on the ways that Compute Canada has met the community's needs and on the areas in which more advanced research computing support would help scholars in the humanities advance their programs of research. To this end, the pages that follow address staffing and training; allocation processes; and sustainability, data management, and long-term storage.

## **Staffing and Training**

The 2014 whitepaper requested a dedicated digital humanities specialist, and the community thoroughly appreciated how quickly Compute Canada acted to meet this request. Community members have been unanimous and vociferous in their praise for John Simpson. He has done an outstanding job communicating the needs of humanities researchers to Compute Canada, doing outreach (as the substantial increase in the number of Compute Canada humanities and social science users in the last year evinces), and encouraging the digital humanities community to access and use the computational resources necessary for their research. As a result of John's significant efforts, members of the digital humanities community feel as though their local Compute Canada representatives better understand their research objectives. The community is

also glad that Megan Meredith-Lobay, with her particular background in humanities computing, has joined the Compute Canada team.

Compute Canada's work to better understand humanities research has been tremendously welcome. We are glad of outreach through the Digital Humanities Summer Institute, as the non-humanities members of Compute Canada's digital humanities team seek to understand our fields' unique research questions and concerns (it is by analogy that we imagine how challenging this must be: akin to sending domain experts in Early Modern literature and culture to learn about the research needs of physicists). The community is also glad of the time and resources that Compute Canada has put into their partnership with Software Carpentry to help faculty and HQP develop their research computing skills.

Currently, the community's concerns about staffing and training stem from the distribution of resources and the communication challenges between Compute Canada staff members who do not have a background in the humanities and humanities researchers who do not have a background in the fields that have traditionally used HPC. Some community members are concerned that interdisciplinary differences occasionally make it difficult to communicate humanities research needs to Compute Canada staff, and that as a result the field of digital humanities is coming to represent every discipline beyond those that have historically worked closely with Compute Canada. There has also been some concern that Compute Canada staff with whom humanities researchers have built an especial rapport may be stretched and overburdened.

We would like to propose a number of potential solutions to these challenges:

- The addition of a social sciences specialist to the Compute Canada complement, since the research questions and computing practices in the social sciences (which often centre on the study of contemporary social relationships) are quite different from those in the humanities (which often focus on the study of the historical visual and textual record).
- A more even geographical distribution of digital humanities personnel with operational expertise. Personnel with humanities expertise would be welcome in Ontario and eastern Canada. Some of the other suggestions offered here (for example, a glossary for the RPP application and a calculator to work out core-year needs and the in-kind value of Compute Canada resources, below) would free up the time of staff members who are currently glossing RPP terms and helping applicants work out in-kind calculations on a one-on-one basis.
- A model for long-term embedded research infrastructure operational support (based on, perhaps, both competition and payment), that would include support and maintenance for software stacks above the operating system level.
- Ongoing support for a tailored version of Software Carpentry that teaches Software Carpentry's current competencies, but recognizes, for example, the material and

computational approaches unique to the humanities. We would be glad to see Software Carpentry feature use cases for the various Software Carpentry tools (command line, git, R, etc.) that reflect the research interests of humanities researchers. Continued development of humanities-oriented curricular materials would be very welcome. Compute Canada might also consider ways that those who take Software Carpentry can be supported in continued learning by, for example, drawing on and integrating with existing training programs, like those pioneered by the DHSI.

## Allocation Processes

The digital humanities community has been glad of the diversification of Compute Canada's allocation streams, as both storage and computing are important in the digital humanities. While, at the moment, most humanities researchers are only taking advantage of Compute Canada's default allocations, we are starting to see a change in the community's needs. The experience of the humanities researchers who have taken on an RPP application (and those of the community who have opted to undertake high performance computing using resources other than Compute Canada's) are illuminating, as they offer insight into the challenges that might hinder further uptake from the humanities community.

We had extensive feedback that included suggestions for improving the RPP and RAC application processes. Some members found that a lengthy application which required forecasting high performance computing needs a year in advance and the lack of a comprehensive glossary of terms (of, for example, what *core years* means to Compute Canada), made the application process unduly onerous. Some researchers who had stored their data with Compute Canada raised concerns about the data loss they encountered due to a lack of redundancy and backup of the Compute Canada systems they were using. Researchers facing these challenges have opted to use Amazon's cloud services or university-specific services in order to either dedicate the time that an RPP or RAC application would otherwise take to research or to ensure redundancy and more robust backup.

The default allocations are excellent, but the increase in the number of RPP and RAC applications in the humanities suggests that the need for storage, transfer, and processing is likely to grow. The following concrete solutions would help meet the needs of the community:

- Better documentation generally of the application forms and processes. This includes streamlining where possible, to make the application process less arduous. We would like better documentation of the way that RPP allocations work (overall a multi-year commitment but with specific allocations for only one year), so that researchers can understand allocations more readily. Clarification of what kind of staffing allocation and support, if any, can be made available through an RPP would be ideal. The community would welcome comprehensive glossing of RPP and RAC application terms in the

Competition Guides, not only to assist applicants who are new to advanced research computing, but also to help applicants communicate the value and necessity of these resources to colleagues, grant assessors, and administrators who are not engaged in advanced research computing. We would also welcome greater support from on-campus Compute Canada personnel at the application stage.

- Closer integration of the RPP and RAC application process with SSHRC, NSERC, and CIHR applications. Closer integration would both cut down on the duplication in the application process and on the number of scholars who are left with the funds for HQP but insufficient computing and storage resources, or sufficient computing and storage resources but no HQP to use them. An online calculator that would help researchers determine their need and articulate in-kind value to funding agencies would be a practical way to solve some of the challenges listed above. We also believe that integration with SSHRC might help with Compute Canada's visibility (as humanities researchers are, as yet, largely unfamiliar with Compute Canada).
- Streamlining of Compute Canada cloud services to make them as easy to use and secure as Amazon's services, to maximize the utility of Compute Canada to the numbers of scholars who have currently opted to have their computing resource needs met elsewhere.
- Setting aside a percentage of Compute Canada's resources for experimental jobs, with an abbreviated application process and use timeline, modeled, perhaps, on Compute Canada's existing fast-track competition.
- Continued improvements to the robustness and performance of the cloud-computing infrastructure, particularly for web-based applications. There have been several issues — such as unreliable bandwidth, data loss, and problems with disk imaging— that haven't helped adoption.

## **Sustainability, Data Management, and Long-Term Storage**

The community has responded with real enthusiasm to Compute Canada's plans for increased storage options and for cloud computing resources that are available to researchers from any institution. CSDH/SCHN's 2014 whitepaper pointed to challenges in creating connections between the foundational infrastructure provided by Compute Canada and its regional partners and the top layer of discipline- and researcher-specific research, where, as outlined in the paper and above, there is often duplication of Compute Canada services and infrastructure purchased through grant funding. Compute Canada's expansion of training opportunities and addition of humanities staff has done much to create that middle layer that bridges the gap between researchers and infrastructure. The community would still welcome software as a service as a way to support research in a way that would bring the top and foundational level together.

Many of the current challenges arise from the short Compute Canada access and allocation windows. Annual renewal for faculty and students is onerous, and, when it comes to hosting tools or platforms, the three-year RPP window is rather short. We recognize that it is not within

Compute Canada's remit to store datasets and tools in perpetuity. However, the process for safely removing one's data and tools for archiving elsewhere remains opaque. If Compute Canada is confirmed as the national provider of infrastructure for advanced research computing, then the lack of an alternative in such cases needs to be addressed as does the question of backup.

These challenges are not insurmountable. We would like to suggest the following possible solutions:

- Coordinating the account renewal for faculty and students is onerous. We would welcome automatic renewal for active users and the users they have sponsored, with a more proactive renewal process only for users whose accounts have become dormant.
- Supporting more software as a service. While the members we consulted were split on the question of Compute Canada's current SaaS (some members suggested that Compute Canada needn't replicate services already offered by, for example Dropbox or Skype, with others suggesting that access to OwnCloud is an excellent on-ramp for faculty whose only experience with and conception of cloud computing is, to date, Dropbox. Others have been glad of the services for researchers in the humanities who have reservations about storing, or often are prohibited from storing, their data on servers in the USA). There are robust and popular humanities tools, such as Voyant and New Radial, which, if backed by Compute Canada's computing power, would be of even greater utility to the community. The community would also be glad of notebook hosting for software like the iPython Notebooks or Databricks, which would cut down on the duplication of effort that it takes scholars to set up local instances of this software.
- Access to Windows-based virtual machines, like the ones created to support SSHRC's Open Data Challenge. Given the extent to which Windows is used in our community, we expect that support for Windows would make Compute Canada services more accessible to a sizable group of potential users.
- Storage and hosting that lasts longer than three years. Knowing, however, that Compute Canada cannot archive data forever, we would welcome clearer roadmaps for exporting our data and migrating hosting from Compute Canada.

We have been glad of this call to reflect on the ways that Compute Canada has worked to meet the humanities community's needs since 2014. The various options for SPARC2 feedback, via surveys, papers, and in-person and virtual meetings have been excellent. We are keen to continue to work with Compute Canada to increase humanities scholars' awareness of and involvement with Compute Canada, as we work to answer humanities research questions together.

March 2016