

High-Performance Computing: An Agenda for the Social Sciences and the Humanities in Canada

by

John Bonnett
Tier II Canada Research Chair in Digital Humanities
Department of History
Brock University
500 Glenridge Avenue
St. Catharines, ON
L2S 3A1
jbonnett@brocku.ca

Abstract: CFI has just announced its financial support for the creation of a national High Performance Computing (HPC) platform, and the formation of a new organization to govern it: *Compute Canada*. When realized, the platform will afford Canadian scholars with more computational power than they have ever enjoyed before. HPC presents rich research possibilities for Canada's social scientists and humanities (SSH) researchers. Our respective research communities, however, are not yet prepared to exploit them. This report outlines two possibilities that HPC presents for SSH research, centering on serious computer games and Massive Multi-User Persistent Worlds. It also contains recommended steps for SSHRC and the SSH research community to take in order to exploit them.

In November 2006, the Canada Foundation for Innovation (CFI) announced its financial support for an initiative to create a national High Performance Computing (HPC) platform, and a new organization to govern it: *Compute Canada*.¹ The funding application was submitted by the seven regional networks that now support HPC in Canada: ACENET²; CLUMEQ³; SCINET⁴; HPCVL⁵; RQCHP⁶; SHARCNET⁷; WESTGRID⁸. Combined with matching funds from provincial and other granting agencies, the grant will infuse \$150 million to support the construction, operation and management of Canada's next generation HPC infrastructure, with four regional networks being the beneficiary of this round of funding: CLUMEQ; RQCHP; SCINET; and WESTGRID. CFI assumes that further funds will be available in three years time to upgrade the infrastructure of the remaining networks, and to enhance the capacity of the national network as a whole. CFI will contribute \$60 million of the proposed \$150 million through its *National Platform Fund* (NPF). The NPF fund was established for two reasons: to support the development of HPC in Canada; and to support the development of knowledge management resources for the humanities and social sciences. With respect to HPC, the NPF is designed to support three aims: to ensure access to internationally competitive HPC resources that will benefit Canadian institutions and researchers; to raise Canada's international profile in HPC; and to provide a stable funding base to support the development of present and future HPC infrastructure. Prior

to 2005, CFI was subjected to multiple applications from multiple consortia as each organization's need to upgrade infrastructure emerged. Members of Canada's research community realized the need for a more stable and coordinated source of funding, and in 2005 c3.ca expressed that need in its Long Range Plan (LRP) for HPC in Canada.⁹ CFI's creation of the NPF fund was facilitated, in part, by c3.ca's LRP.

In its call for proposals for the NPF, CFI made clear its position that all future HPC research must be undertaken through the proposed national platform. No independent HPC initiatives will be funded. The proposed infrastructure must meet the needs of all disciplines within Canada's research community, from particle physics to history and computing. Accordingly, *Compute Canada* approached select members of Canada's human science community to determine their needs. And accordingly, now that *Compute Canada*'s application has received a favourable decision from CFI, Canada's social science and humanities scholars (SSH) should consider the implications that the proposed HPC platform will present for their disciplines. Will *Compute Canada*'s proposal – with respect to infrastructure, with respect to operation, with respect to governance – enable SSH researchers to conduct leading edge research devoted to computationally challenging problems?

As currently conceived, I do not believe the *Compute Canada* proposal meets that mandate for human science researchers. I further suggest that *Compute Canada* could not have met that mandate, even if it had enjoyed an extended period to assess SSH scholars' needs, which it did not. Canadian social science and humanities scholars will soon enjoy access to unprecedented levels of computational power. The unfortunate fact, however, is that most scholars in our respective communities have no idea how to exploit that power to support their research. And even if they did, SSH researchers, unlike their colleagues in the natural and medical sciences, do not possess the computational expertise necessary to use the proposed network, or specify their requirements for it.

If Canada's human science community is to make meaningful use of the country's emerging HPC infrastructure, I suggest there a number of prerequisite steps it must take:

- It must first define important problems requiring HPC levels of computation.
- It must define its infrastructure needs to support its research, and the forms of access it will require to that infrastructure.
- It must discern how its infrastructural and operational requirements differ from research communities in the natural and medical sciences.
- It must determine what support it will require – in the form of instruction, in the form of programming – from personnel affiliated with the national HPC platform
- It must communicate those needs to CFI and, if realized, *Compute Canada*.

SSHRC, I suggest, can and must take a leading role in encouraging scholars' migration to High Performance Computing.

My arguments regarding the *Compute Canada* proposal, and the steps Canada's SSH community must take to engage in HPC, rest on a specific conception of what constitutes a "leading edge" problem. It is not a conception that will be shared by all social science and humanities scholars. Currently, Canada's HPC networks are configured to support the processing and mining of large quantities of data, and the creation of highly complex simulations in domains ranging from astrophysics to

proteomics. Projects are run on a queued batch basis, meaning you submit your project, the HPC networks computes it, and then the results are returned to you. There are human science scholars who can and do operate comfortably in such an HPC regime. Economists, for example, are currently deploying micro databases and econometric models in high performance computing networks to assess the probable impact of social policies. In the domain of digital humanities, HPC could potentially be used to support text analysis and text mining.

Doubtless, the projects described above can and should be characterized as leading edge. However, to delimit human science high performance computing to these initiatives, and no other, would in my opinion be a serious failure of imagination. There are other possibilities that Canadian researchers can pursue, and should pursue due to the potential social and economic benefits they would confer. Realization of such a research agenda will require *Compute Canada* to change the form of access it provides researchers, from queued batch processing to interaction. It will further require the organization to dedicate parts of Canada's HPC infrastructure to meet the specific needs of Canada's human science research community. Instead of simply running a program, I suggest social science and humanities researchers will eventually undertake research in which human subjects interact with simulations or persistent virtual environments in ways that are useful for research. Two examples will clarify my point.

One possibility – of potential interest to the social sciences – would be to devote HPC infrastructure to projects that combine large-scale agent-based simulations with serious games. Among other things, agent-based simulations are currently being deployed to assess the potential epidemiology of pandemics. Recently, Chris Barrett and his team at Los Alamos National Laboratory created a virtual Portland comprised of one million agents to determine the trajectory of a smallpox pandemic.¹⁰ The purpose of such HPC simulations is to determine the potential hotspots for transmission, to enable policy-makers to contain and mitigate the transmission of smallpox and other contagions.

This is important work. I suggest, however, that the value of such simulations could be enhanced if they were combined with serious games. Serious games are games that appropriate the technologies and expressive forms of computer games for purposes ranging from military training to education and public safety. The research purpose of combining serious games and HPC simulations would be to assess the strengths and weaknesses of the operational constructs Canada's three levels of government deploy to contain and mitigate public safety and environmental threats. In essence, when jurisdictions face an environmental or public safety threat, they deploy agencies, equipment and individuals to meet that threat. Serious games, combined with training and exercising, would enable policy-makers to determine how well current or proposed operational constructs are functioning. Who is cooperating? Who is not? What policies are working? Which are not? The public benefit that would result from the combination of IT, HPC and training and exercising cannot be emphasized too highly. Emergency responders in the aftermath of 9/11 indicated that preparation for the Y2K incident was invaluable in assisting them in their response and recovery to the World Trade Center crisis.¹¹

A second potential domain of research – of potential interest to the humanities – would centre on the effective deployment of massive multi-user persistent worlds (MMPWs). MMPWs are on-line virtual environments that enable hundreds if not

thousands of users to interact in real-time. Their initial domain of application has been entertainment games, but non-game applications ranging from communications, training, distributed collaboration, predictive modeling, and education have also been identified.¹² The significance of MMPWs is two-fold. Culturally, its advent is as important as the emergence of the codex in ancient Rome. MMPWs, like computer supported collaborative environments, are novel platforms supporting the generation, analysis, storage, and distribution of information. Their effective implementation will require the generation and testing of novel conventions for narration, expression and documentation. MMPWs, like film at the start of the 20th century, will require the development of a “grammar” specific to the medium of virtual environments. Humanities scholars are well placed to contribute to such a practice-based research agenda. They have a deep knowledge of extant conventions in domains ranging from art, to literature and film, conventions that could be adapted. And more to the point, they have undertaken this type of research before. In the domain of documentation, for example, the footnote emerged as the product of innovation, testing and debate in the 17th century.¹³

The MMPW is further significant for its potential economic impact. This form of argument is one that social science and humanities scholars generally prefer to avoid, largely because many feel that the value of research should never be subordinated to economic considerations. Research, particularly in the human sciences, should be judged by other criteria, its contribution to knowledge, and its potential cultural, social or political impact. I understand the position. In some cases, I even sympathize with it. But with respect to the future deployment of HPC in the human sciences, I suggest such a position is counter-productive. If SSH researchers move into the domain of HPC along the lines I am describing, they will be asking a great deal from CFI and *Compute Canada*. In essence, they will be asking for clusters specifically dedicated to SSH projects. Such a request will be necessary because researchers will require access to HPC either on a protracted basis (one, two, or three days) – in the case of serious games – or a persistent basis (months, potentially years), in the case of MMPWs. For *Compute Canada* and CFI to agree to such a request, the SSH community must provide arguments that will resonate with policy-makers. Economic arguments, in my estimation, will provide the most persuasive rationale for CFI and *Compute Canada* to allocate HPC resources for the purposes we designate.

This point can be elaborated by considering the impact of Canada’s cultural sector generally, and the potential economic impact of MMPWs particularly. In 2001, Statistics Canada reported that Canada’s arts and cultural sectors contributed some \$26 Billion dollars to the country’s gross domestic product.¹⁴ On a more concrete level, the producers of the science fiction series *Stargate* have estimated that the franchise has generated some US\$ 500 million for the British Columbia economy in the past 10 years.¹⁵ With respect to the potential economic impact of MMPWs, in 2005 the game *World of Warcraft* generated some US\$250 million of revenue.¹⁶ These figures to my mind present five important implications to Canada’s human science community and SSHRC:

- First, Canada’s cultural industries demonstrably make an important contribution to Canada’s economy.

- Two, cultural industries are not static things. Their constituents evolve. New forms of culture emerge, and these forms have economic implications.
- Three, the effective deployment of new cultural platforms and formalisms requires research to ensure that their full potential for expression and social dissemination is realized.
- Fourth, our objects of research, particularly MMPWs, are by nature persistent, and accordingly will require the long-term dedication of HPC resources.
- Fifth, given the actual and potential impact of Canada's cultural and digital cultural industries, Canada's SSH community and SSHRC have grounds to request dedicated clusters that permit interaction, as opposed to batch processing of research projects.

This conclusion raises the question as to what SSHRC should do now and in near and far future with respect to facilitating the integration of HPC into the research agendas of the disciplines it represents. I have no specific policy recommendations. What I can offer is three principles that I believe should govern the formation of future policy.

Principle One: SSHRC and Canada's social science and humanities research communities should pursue the strategic objective of obtaining dedicated clusters in future iterations of Canada's HPC infrastructure.

For the reasons specified above, SSH researchers will require the capacity to interact on a sustained basis with the virtual environments and simulations they are using to support their research. They will further require dedicated clusters within Canada's HPC infrastructure to support that research. However, I do not believe the acquisition of such clusters to be an achievable aim for this round of funding. There are two reasons why. One, the community of humanities and social science researchers is currently too small to merit allocation of one or more clusters. Two, *Compute Canada's* proposal has already been approved for funding. It was based on a certain understanding of the needs of our various research communities. Our respective communities no longer have the time or opportunity to change that understanding. Our efforts are better dedicated toward preparing for future iterations of Canada's HPC infrastructure.

Principle Two: To realize the aspiration of increased access to HPC, SSHRC should develop a deeper working relationship with CFI and *Compute Canada*.

Rationale: At this stage, there is little reason to expect more from CFI and *Compute Canada*. If HPC is to be integrated into SSH research and practice, I am suggesting the human science community and SSHRC, on the one hand, and CFI and *Compute Canada*, on the other, must jointly make that integration a long term strategic goal to be realized in ten to twenty years time.

If SSHRC means to encourage the incorporation of HPC into social science and humanities research, it must deepen its relationship with CFI at all levels, and develop a relationship with *Compute Canada* so that there is an understanding in both organizations of the present and future research needs of our respective communities.

SSHRC and SSH scholars, for their part, must present CFI and *Compute Canada* with concrete research initiatives along the lines I have described. Otherwise, neither organization will have an incentive to adjust the configuration and operation of Canada's HPC infrastructure.

Principle Three: SSHRC, CFI and *Compute Canada* should take individual and collective steps to encourage the next generation of SSH researchers to incorporate HPC into their research.

Rationale: Such encouragement, I believe, will not come from established scholars in the disciplines represented by SSHRC. Most have a surface understanding, if any, of high performance computing. Most would not likely know how to exploit it even if they did have a deeper understanding. The potential for social science and humanities HPC rests with today's graduate students. Realization of that potential will only come if they are exposed to the opportunities that HPC will afford.

¹ *Compute Canada* will replace c3.ca, a national organization that was founded in 1997 to advocate high-performance computing in Canada. The new organization will function both as an advocacy group, and as a governing body to ensure the proposed platform functions as a national infrastructure. Currently, c3.ca comprises 50 member institutions, and represents the HPC interests of thousands of Canadian researchers.

As conceived, *Compute Canada* will be comprised of researchers, the seven consortia, and universities
² Atlantic Computational Excellence Network. **Members:** Dalhousie U., Memorial U., Mount Allison U., St. Francis Xavier U., St. Mary's U., U. of New Brunswick. U. of Prince Edward Island. *Soon to join:* Acadia U., Cape Breton U.

³ Consortium Laval UQAM McGill and Eastern Quebec for High Performance Computing. **Members:** McGill U., U. Laval, UQAM, and all others branches and institutes of l'Universite du Quebec: UQAC. UQTR, UQAR, UQO, UQAT, ETS, ENAP and INKS.

⁴ Réseau québécois de calcul de haute performance. **Members:** Bishop's U., Concordia U., Ecole Polytechnique, U. de Montreal, U. de Sherbrooke.

⁵ High Performance Computing Virtual Laboratory. **Members:** Carleton U., Loyalist College, Queen's U., Royal Military College, Ryerson U., Seneca College, U. of Ottawa.

⁶ SciNet. **Member:** University of Toronto.

⁷ Shared Hierarchical Academic Research Computing Network. **Members:** Brock U., Fanshawe College, U. of Guelph, Lakehead U., Laurentian U., Sir Wilfrid Laurier U., McMaster U., Ontario College of Art and Design, U. Ontario Institute of Technology, Sheridan College, Trent U., U. of Waterloo, U. of Western Ontario, U. of Windsor, York U.

⁸ Western Canada Research Grid. **Members:** Athabasca U., Brandon U., Simon Fraser U., U. of Alberta, U. of British Columbia, U. of Calgary, U. of Lethbridge. U. of Manitoba, U. of Northern British Columbia, U. of Regina, U. of Saskatchewan, U. of Victoria, U. of Winnipeg.

⁹ The Long Range Plan, titled *Engines of Discovery: The 21st Century Revolution*, is available on-line at: www.c3.ca/LRP [October 10, 2006]

¹⁰ Chris L. Barrett, Stephen G. Eubank and James P. Smith. "If Smallpox Strikes Portland... 'Episims' unleashes virtual plagues in real cities to see how social networks spread disease. That knowledge might help stop epidemics." In *Scientific American*. 292(3): 54-61. March 2005. Available on-line at:

<http://www.sciam.com/article.cfm?chanID=sa006&colID=1&articleID=000BBC08-CEA3-1213-8EA383414B7FFE9F> [October 11, 2006].

¹¹ One recent example of a social science contribution toward improving the operational effectiveness of policy-makers and first responders is Sharon S. Dawes, Anthony M. Cresswell, and Bruce B. Cahan. "Learning from Crisis: Lessons in Human and Information Infrastructure From the World Trade Center Response." In *Social Science Computer Review*. 22(1): 52-66. Spring 2004.

¹² Robert Gehorsam. "The Coming Revolution in Massively Multiuser Persistent Worlds." In *IEEE Computer*. 36(4): 93-95. April 2003.

¹³ Anthony Grafton. *The Footnote: A Curious History*. Cambridge, MA: Harvard University Press, 1997.

¹⁴ The Canada Council for the Arts. "Creating Jobs and economic impact." At <http://www.canadacouncil.ca/aboutus/advocacy/hj127323047009843770.htm> [October 11, 2006]. It is also worth considering that the Canada Council is the source for this statistic. Its appeal to economic arguments suggests that economic and artistic interests are not always at odds. A similar argument could be made that there is not always a conflict between the research interests of SSH scholars and research initiatives that would serve the economic interest of Canada.

¹⁵ Mary McNamara. "Stargate 200: Science-Fiction Series 'SG-1' Is Cable's First to Reach Historic Milestone." In *Multichannel News: A Variety Group Publication*. At <http://www.multichannel.com/article/CA6332083.html> [October 11, 2006].

¹⁶ Seth Schiesel. "World of Warcraft Keeps Growing, Even as Players Test Its Limits." In *New York Times*. February 10, 2005. Available on-line at: <http://www.nytimes.com/2005/02/10/technology/circuits/10warr.html?ei=5090&en=ba633bc16b682182&ex=1265691600&partner=rssuserland&pagewanted=all&position=> [October 11, 2006].