

# Mind the Gap

## A Multidisciplinary workshop bridging the gap between High Performance Computing and the Humanities

**Draft Report 2.0** prepared by Geoffrey Rockwell and Megan Meredith-Lobay  
with help from Jeff Antoniuk, and Jon Johansson

### 0. Executive Summary

There is a gap between research in the Humanities and Canadian high-performance computing (HPC) facilities, but it is not what we thought it was. We used to think humanists didn't need supercomputing - they were happy with a wordprocessor, email and the Web. Now it is clear that humanists have large multimedia datasets and big questions to ask of the history of human culture. Then we used to think the gap was primarily between facilities set up for queued batch programs and practices in the Humanities of asking questions repeatedly of "always-on" web services. Though there is still some truth to that gap, many HPC facilities have begun to support "portal" or "cloud" facilities that are always-on and can thus support Humanities practices. The gap now is really one of research culture and support. On the one hand we have to find ways of training and preparing humanities research teams to be able to imagine using existing HPC facilities, and on the other we have to develop the ability of HPC consortia to be able to reach out and support humanists.

This Report provides:

1. **Background** information about the workshop.
2. Discussion of the **achievements** of the workshop.
3. **Recommendations** for bridging the gap.
4. **Next Steps** for HPC and the Humanities.
5. How to run your own **Mind the Gap**.
6. **Links**.

It became clear that there is an opportunity to define the HPC needs of digital humanists in a way that can be used to develop collaborations with organizations like Compute Canada. These needs have been articulated as recommendations, but they really are really a set of ideas for collaboration, not an expectation of support. The challenge is how to work with organizations from our universities to SSHRC to develop appropriate support. These recommendations come out of our experience with what actually worked.

1. Humanists need **data acquisition infrastructure** integrated into the research pipeline. Infrastructure for acquiring research data in digital form should be integrated into computing facilities in partnership with other units like the Library.

2. Humanities HPC facilities should offer **"always on" computing services** that support research practices in the Humanities that typically are not well supported by batch models.
3. Humanities HPC facilities should support **programming languages commonly used in the digital Humanities** like Java and PHP. A library of optimized common operations used in Humanities HPC applications could be developed to compensate for the inefficiencies of the languages humanists typically work with.
4. An **easy-to-use visual shell** or interface could be adapted for humanists to be able to queue common jobs.
5. **Visualization** facilities with appropriate support should be developed to support researchers across the arts and humanities.
6. **Group Conferencing** facilities should be adapted to be easily booked, set up and used by faculty in the Humanities.
7. HPC facilities could **partner with archives** to provide large-scale distributed text/data repositories with off-site redundancy. Such partnerships need to be able to provide security for protected materials.
8. Given the importance of experimental dissemination in the Humanities, HPC facilities should consider setting up **dissemination sandboxes**.
9. Digital humanists should take the lead in developing **Training** packages for their local facilities. **Common training materials** could be developed by the digital humanities community once for adaptation by others.
10. The **Mind the Gap model** should be adapted so that it can be offered at other facilities to enable teams across the country to take advantage of Compute Canada.
11. HPC facilities should have **staff familiar with Humanities research** practices who can provide outreach, advise and support researchers new to HPC. Digital humanists can server as **mentors**.
12. Humanists should get involved appropriately in the **development and management** of HPC facilities.
13. SSHRC and other funding bodies should develop **competitive programs for supporting** HPC projects in the humanities and social sciences.
14. HPC facilities should work with their participating universities to develop competitive **Faculty Fellowships** and **Graduate Fellowships** that provide the support needed to develop a project.
15. The Humanities community have to **make the research and business case** for HPC support. It is our responsibility to show the need and to work with others to develop facilities.

## 1. Background to the Workshop

**Mind the Gap (MtG)** was a one week workshop run at the University of Alberta from May 10th to May 14th. The workshop was designed to:

- To **advance** real research projects with intensive support from HPC personnel so that they are able to continue working with WestGrid and apply for funding.

- To develop an interdisciplinary **community** of researchers, both at the University of Alberta and across Canada, looking at the application of HPC techniques to Humanities research. To encourage the community to learn from each other and to support each other.
- To define an **agenda** of interesting Humanities research problems that are tractable using the HPC facilities at hand, notably WestGrid.
- To **train** researchers in the Humanities in the skills needed to use HPC facilities like WestGrid.
- To **engage** WestGrid staff and computer scientists with Humanities problems in order to develop opportunities for interdisciplinary research.
- To develop **ideas** for next steps that can develop the research community and strengthen ties with the HPC community.

We proposed the workshop as a way of bridging the research cultures of the Humanities and HPC. The two cultures are currently too far apart for humanists to easily imagine using HPC methods, technologies or facilities like WestGrid, let alone applying for grants to do HPC research. HPC facilities are typically set up for batch-style computing where very large problems are programmed so they can be queued up and submitted for processing by compute clusters. Humanists by contrast tend to use computing to publish electronic texts and other forms of evidence to the web where the resources are “always-on” and searchable for reading. The first step therefore to using HPC methods and facilities for Humanities research is bridging the cultural differences and learning from each other. Further, humanists typically don’t have the preparation or access to programmers to use HPC clusters or visualization tools. MtG therefore offered a workshop designed to train Humanities researchers so they can collaboratively prototype research problems using local facilities. It was a seed project that brought Humanities researchers together from across Canada (and even the US and UK) to develop prototypes. The workshop had the further advantage of starting a dialogue between researchers in the Humanities, researchers from computing science, and staff with supporting HPC.

Mind the Gap was organized as an "unconference" or unworkshop culminating in presentations from research teams that were supported to prototype real projects. Instead of having preplanned training sessions, MtG was organized around lots of team time. The idea was not to (only) talk about HPC and the Humanities, but to "do it". In order to "do it" we invited teams that had the right mix of programmers and researchers. The invited teams were then set up with access to WestGrid HPC facilities and provided support all week in a flexible meeting space. The support came from HPC staff from AICT, WestGrid and Sharcnet. Woven around the time for teams were invited talks including two talks by invited speakers from outside Canada. The program can be seen here:

MONDAY, May 10	TUESDAY, May 11	WEDNESDAY, May 12	THURSDAY, May 13	FRIDAY, May 14
9:30 - Introductions and Opening Remarks from Geoffrey Rockwell and Paul Lu. Special	9:30 - Invited talk from Robyn Taylor	9:30 - Technical Issues and/or Tour of AMMI	9:30 - Invited talk: Stephen Ramsay:	9:00 - HPC and the Humanities: Introduction to

welcomes from Associate VP Research, Dr. Renee Elio and Associate Dean for Graduate Studies. Faculty of Arts, Heather Zwicker.	"Exploring Human computer Interaction through Performance Practice"	lab with our CAVE	"Knowing it When You See It: Humanistic Inquiry in the Age of Big data"	opportunities with Compute Canada and CANARIE by Jonathan Schaeffer followed by discussion of agenda for research
10:30 - Coffee Break	10:30 - Coffee Break	10:30 - Coffee Break	11:00 - Coffee Break	10:30 - Coffee Break
11:00 - Introduction to HPC at Alberta: Masao Fujinaga will introduce facilities and how to use them.	11:00 - Team Time	11:00 - Team Time	11:30 - Team Time	11:00 - HPC and the Humanities - Ongoing discussion on HPC and the Humanities
12:30 - Lunch	1:00 - Lunch	1:00 - Lunch	1:00 - Lunch	1:00 - Lunch and good-byes
2:00 - Team Time	2:30 - Invited Talk: Patrick Juola: "Computers, Conjectures, and Creativity"	2:30 - Invited Talk: Paul Lu: "Cloud Computing and HPC"	2:00 - GIS for the Humanities, Megan Meredith-Lobay	
4:00 - Break	4:00 - Break	3:30 - Break	2:30 - Team Time	
4:30 - General Discussion	4:30 - General Discussion	4:00 - Team Time	3:30 - Break	
			4:00 Presentations from Teams	
			6:30 - Workshop	
			Dinner at the Faculty Club, University of Alberta	

#### Final Program Including UnConference Additions

In addition to the events originally planned we also set up new events as discussions progressed. An example would be the Thursday 2:00 pm presentation on GIS, which was organized when a number of participants decided this was important.

The morning of the final day was set aside for reflection on HPC and the humanities. It is from that discussion that the ideas and recommendations came.

#### **1.1 Teams**

Mind the Gap took two approaches to identifying teams. Due to funding MtG had no funding for teams to travel to Edmonton. The majority of the teams therefore were made up of researchers at the University of Alberta or teams that had some researchers at U of A. These teams were identified and invited from projects at the U of A that were clearly ready and had the resources to take advantage of the opportunity. This typically meant projects with programmers, well defined research goals that fit, and funding to continue the work.

Mind the Gap also wanted to involve teams from at least one other consortia so we worked with SHARCnet. SHARCnet ran a competition and then funded two teams that were supported by a staff person who also came. We also had a team that included the two invited speakers from the US so they could participant beyond just giving a paper.

#### **1.2 Mind the Gap Technical Organization**



Checkers Cluster, WestGrid

Participants of the Mind the Gap workshop had two nodes reserved on the WestGrid machine "Checkers" at the University of Alberta for their use during the workshop. Each node contains eight CPU cores. The initial workflow we envisioned was the standard one we see with our numerical users: develop a program using tools on the head node, compile and run small test jobs on the head node and finally submit larger jobs to run on the cluster through the batch scheduling system. This model worked well for some of the participants, and in fact the bulk of the CPU time used by participants in the workshop was utilized through the regular batch system. People who were relying on Hadoop needed to run their programs directly on the reserved nodes, so some flexibility in access to the nodes turned out to be important.

The venue chosen for this event utilizes wireless network access only and required authentication using UofA computing IDs. The WestGrid computers are accessed through a separate set of IDs. Initial setup required the following:

- Temporary access to the University Wireless Service
- Temporary IDs for off campus participants set up through AICT
- IDs on HPC resources. In this case, the organizer created a project to obtain his WestGrid ID and workshop participants joined that project
- Teams requested software packages to be installed and configured on the HPC system. Some packages were fairly standard; some like Hadoop were new to us and team members assisted with installation and configuration. Hadoop was configured to utilize only a single node (8 cores) at a time because it wasn't clear how to control Hadoop through our batch schedule. Software included: gcc, Python 2.6 or higher, Java, MySQL, Lucene, Hadoop, Eclipse, Ant and subversion

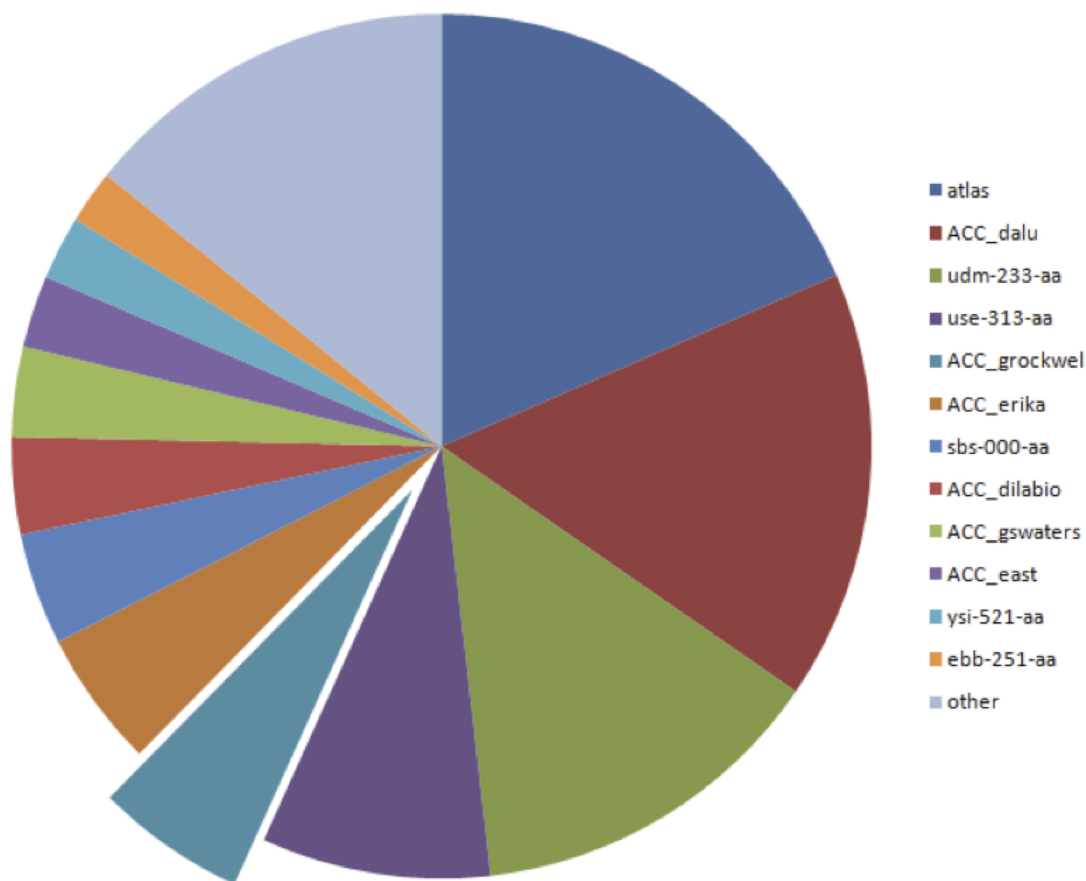
The Monday morning session of the [Mind the Gap workshop](#) was presented by Masao Fujinaga who discussed the facilities available at the University of Alberta, and the bulk of the time was spent having participants:

- Connect to the wireless network
- Download and install applications: PUTTY, WinSCP, XMing
- Connect to the HPC machines
- Edit files
- Submit small batch jobs to the system

This session was very helpful in getting everyone up to speed with using the system. Support from AICT for the rest of the week involved one or two people being present to help with sporadic connection problems and preparation of job submission scripts.

### 1.3 Checkers Usage for the Week

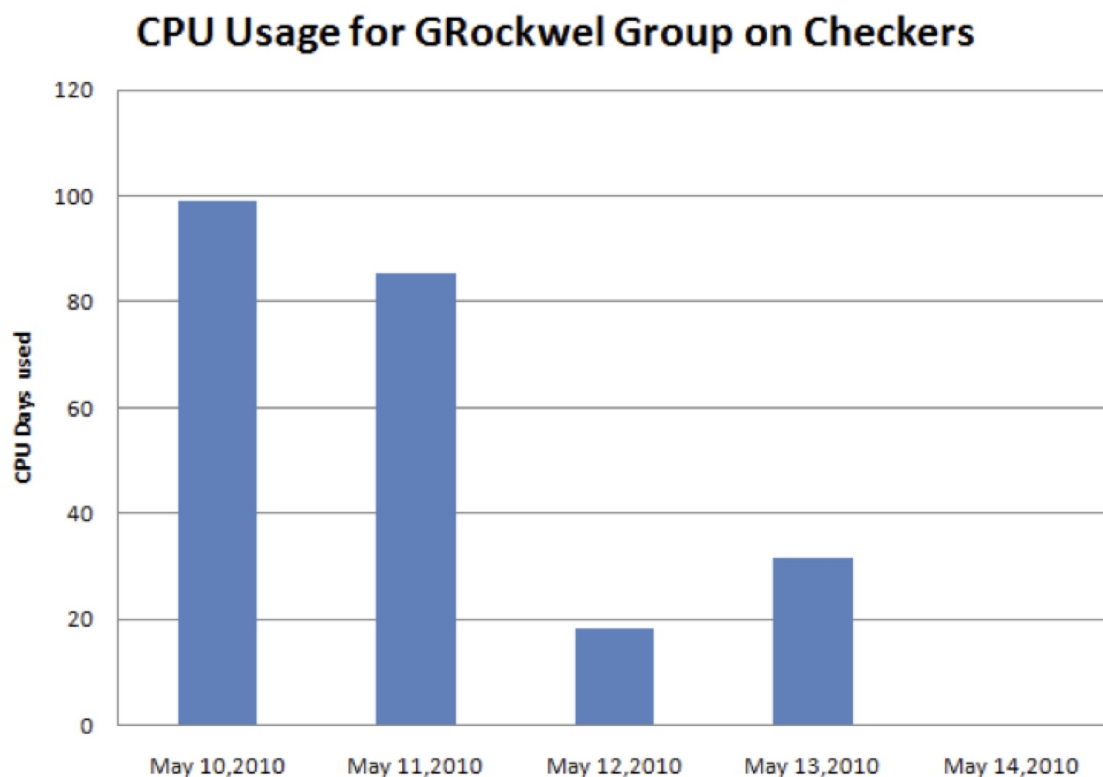
During the week of the workshop some participants made significant use of checkers as shown in the following figure, depicting CPU usage for May 10 to 13 inclusive. The blue exploded wedge compares usage by the 'grockwel' group of 58.5 CPUs to usage of the rest of checkers.



Usage by Mind the Gap Participants (under "grockwel") Compared

The group used a total of 234 CPU days of processing on the checkers cluster during the week i.e. if the jobs had run on a single processor they would have taken 234 days. These are only the jobs that ran under control of the batch scheduler on checkers; the people running hadoop were

running interactively so we don't have statistics for them. Daily usage through the batch scheduler is shown below:



Clearly workshop participants came prepared to utilize HPC resources.

## 1.4 Collaborative Organization

Many organizations supported this in different ways. The workshop was a collaboration among different organizations that should be recognized:

- The office of the Vice President Research provided financial support for this workshop.
- WestGrid provided access to their facilities and set aside resources for this workshop. The idea for the workshop came from conversations between Geoffrey Rockwell and Paul Lu, the WestGrid executive member from the University of Alberta. Lu was on the organizing committee and participated in the workshop along with his graduate students.
- Academic Information and Communication Technologies (AICT) provided significant staff support to planning MtG, setting it up, and supporting people at the workshop. There was always at least one staff person who could help and many times there were more staff. Ron Senda and Jon Johannsen were part of the organizing committee from the beginning.
- SHARCnet sponsored two teams. We wanted this to not just be an Alberta event, but to reach out across Canada and SHARCnet stepped up to supporting researchers from Ontario. This was coordinated with Hugh Couchman. Dave McCaughan from SHARCnet came to the workshop and provided valuable support.



- The Faculty of Arts was also a supporter, especially through the Arts Resource Centre of the Faculty. Megan Meredith-Lobay of ARC was on the organizing committee and did much of the work around the web site, organizing the space, and organizing events.
- The workshop was initiated and led by Geoffrey Rockwell, Director of the [Canadian Institute for Research in Computing and the Arts](#).

## 2. Achievements of the Workshop

The Mind the Gap Workshop achieved its goal of bridging the gap between HPC and the Humanities, but in the process achieved much more in the way of on the spot results and in the development of a long-term strategy for integrating HPC into Humanities Research. The Teams were invited to the Mind the Gap workshop in order to develop prototypes for DH projects which could utilize HPC resources. Each team achieved their goals and were able, by the end of the week, to produce meaningful results. In the process, teams merged and split, shared data and results, and collaborated to produce new results not originally part of any original project description. The process of sharing data and expertise was a benefit not only to the Humanists present, but also to the members of the Computing Science and WestGrid communities who were the primary programmers for many of the projects. Conversations between CS members and Humanists both on the same team and between teams allowed the Computer Scientists to envision the Humanities problems as computing problems at a new level. In addition, the act of building prototypes, testing, and presenting allowed an unprecedented level of direct cooperation between the different fields.

The format of the workshop meant that problems were tackled at the most practical level, applications built, data tested and run at extraordinary speeds, visualized, and shared with a community of the willing. In the process of doing, the workshop became a chance for contextualized training for all graduate students, faculty, and staff present. This training and learning did not flow only one way over the course of the week, but, as mentioned above, became more of a conversation between the Humanities and CS communities about what else might be possible. This conversation led to the development of a concrete set of recommendations and the articulation of needs which serve Humanities scholars who wish to engage with HPC resources, but also provide the HPC community, WestGrid and SHARCnet, with new ways of engaging with a new constituency.

In sum, some of the achievements of the workshop were:

- Teams actually developed prototypes that ran on a cluster.
- Teams learned from each other.
- The workshop provided time for sustained conversations between humanists and computer scientists on teams.
- WestGrid and AICT staff got a chance to engage a new constituency and learn about this constituency.
- It provided contextual training for graduate students, faculty and staff.

- On the last day we had a sustained discussion of HPC and the Humanities that included developing a list of recommendations.

### 3. Recommendations for Bridging the Gap

#### 3.1 General Observations

Other workshops like the [Digital Humanities and High Performance Computing](#) workshop put on by SHARCnet have explored the main difference between the research computing culture of the Humanities and that of typical HPC facilities. The major difference is between facilities set up to offer queued access to large clusters on which batch jobs are run and practices in the Humanities where researchers are used to iteratively asking questions of evidence. This workshop went beyond discussion and created a context where projects could prototype and try using HPC facilities set up in different ways. One of the general insights is that there is a significant incremental cost for Humanists experimenting with HPC techniques. The time it takes for a researcher unfamiliar with HPC simply to understand enough to be able to articulate their research in a way that is tractable discourages most from experimenting. Why invest in risky learning to discover that HPC is not really for you? Why invest significant time just to get to the point where you can start a project when you could have written a paper or two in that time? The key, therefore, to the engagement of the Humanities is to reduce the risk and the incremental costs. The solution, as we discovered in our discussions, has to do less with technology and more to do with people and their support, something that Compute Canada is poorly funded to provide. Mind the Gap, of course, was designed to help offset the cost of preparation by providing intensive support so that projects could create a prototype that would in turn let them make decisions and prepare grant proposals.

Mind the Gap also had the morning of the final day set aside for discussion on how to develop HPC in the Humanities. The general outlines of how humanists can be engaged include:

- Support the complete research pipeline (3.2),
- Provide customized outreach, training, and prototyping opportunities designed to help new Humanities projects efficiently (3.3),
- Involve humanists in the management and decisions of HPC facilities so they have a say and a stake (3.4), and
- Where there are specialized research needs, work with the digital Humanities centres and other units to develop those (3.5).

#### 3.2 Support the Complete Research Pipeline

In his presentation at MtG, Johnathan Schaeffer argued that HPC facilities should support the complete research pipeline which means supporting more than just the computing. We find this a useful way to think about how support for Humanities HPC research could be structured.

**R.1 Recommendation:** Humanists need **data acquisition infrastructure** integrated into the

research pipeline. Infrastructure for acquiring research data in digital form should be integrated into computing facilities in partnership with other units like the Library.

One of the major transformations of this epoch is the translation of the human record into digital form so that it can be studied by computing methods including HPC and visualization. Digital humanists and librarians are leading the creation of large-scale scholarly electronic archives. This may be one of the most important contributions to the Humanities of our generation. For this reason, and because we believe HPC consortia should support the full pipeline, it makes sense to consider acquiring infrastructure for digitizing large collections. HPC consortia should consider partnering with other units involved in digitization to make available the high-end tools that can't be got with research grants, like book scanners and motion capture systems.

**R.2 Recommendation:** Humanities HPC facilities should offer **"always on" computing services** that support research practices in the Humanities that typically are not well supported by batch models.

As has been noted before, one of the greatest differences between the needs of Humanists and the way many HPC facilities are configured has to do with the priority given to a batch-queue model of allocating computing. Many digital Humanities projects require the ability to launch interactive computational jobs on HPC clusters, often from a web server that provides an accessible interface to the researcher. For this reason we recommend that HPC facilities consider offering "cloud" or "portal" types services where projects that need to be "always-on" can run. Paul Lu, in his talk, suggested that it might prove more cost-effective in the long run to provide access to a commercial service like Amazon EC2+S3. Such a "cloud" service could be set up by Compute Canada at one facility and run for humanists across the country. Both SHARCnet and WestGrid are experimenting with such services and their experience could be scaled up to a national service.

At the same time, it should be mentioned that about a third of the computation required during MtG did fit the classic definition of HPC. Text processing can be structured so as to take advantage of queued HPC. For example projects can use a "traditional" cluster to index a large text collection so it could then be accessed through a web server.

**R.3 Recommendation:** Humanities HPC facilities should support **programming languages commonly used in the digital Humanities** like Java and PHP. A library of optimized common operations used in Humanities HPC applications could be developed to compensate for the inefficiencies of the languages humanists typically work with.

Despite the fact that languages like Java are inefficient, the major cost and barrier to HPC usage for humanists is the cost of programming for HPC facilities. Therefore we recommend that facilities be prepared to support projects programmed in languages like PHP and Java.

One way to offset the inefficiencies of Java is to develop a library of the common operations like tokenization that are used in Humanities projects. Then projects could use Java, but call more optimized operations, to be better users of the facilities.

**R.4 Recommendation:** An **easy-to-use visual shell** or interface could be adapted for humanists to be able to queue common jobs.

Such a visual shell might be developed as a web interface with appropriate help that allows people to queue certain common processes like the indexing of large text corpora. A visual shell could be designed so that a programmer could set up the processes that the research uses and then not have to intervene as the humanists loaded new datasets, indexed them and ran queries. Examples of such shells exist and could be adapted.

**R.5 Recommendation: Visualization** facilities with appropriate support should be developed to support researchers across the arts and humanities.

Researchers in the arts and Humanities are experimenting extensively with visualization. There is the opportunity for HCP consortia to easily extend their current visualization facilities to support text and media visualization along with new media research in the arts and Humanities. Visualization is one way for HPC facilities to involve researchers in the arts and Humanities.

**R.6 Recommendation: Group Conferencing** facilities should be adapted to be easily booked, set up and used by faculty in the Humanities.

Faculty are also now familiar with conferencing tools like Skype and using such tools increasingly for research collaboration. Many HPC Consortia have Access Grid facilities that, with some thought and support, could be made easy enough for a humanist to book, walk in, and use for group to group meetings. This is not a cost issue so much as an outreach and support issue. In our experience the technology is there, but it is not structured so as to be an easy service to use. The opportunity is to reach out to any research group that needs to find a room where up to 5 researchers can conference with remote sites. Given the coverage of Compute Canada it should be possible for it to offer a service from any research university to any other.

**R.7 Recommendation:** HPC facilities could **partner with archives** to provide large-scale distributed text/data repositories with off-site redundancy. Such partnerships need to be able to provide security for protected materials.

Central to the Humanities is the preservation, editing, enrichment and study of the human record be it textual or in other media. Digitizing our histories and literatures so that it can be made available is something that touches all Canadians. Whether it is children discovering their heritage or genealogists researching their family, the task of providing long-term trusted repositories with off-site backup is one that HPC consortia could play a role in. Many HPC consortia already have the physical storage and storage at multiple sites. What HPC Consortia don't provide (and shouldn't necessarily have to) is the archiving function, but consortia could partner with libraries, projects, and other units that do have experience with archiving. It should also be mentioned that many data-rich projects need to be able to guarantee that data is secure. Some materials like confidential survey data may be embargoed or have strict access protocols.

**R.8 Recommendation:** Given the importance of experimental dissemination in the Humanities, HPC facilities should consider setting up **dissemination sandboxes**.

For many artists and humanists the creating, writing and publishing is research itself, not just the dissemination of research. HPC facilities should consider providing sandboxes for experimental online projects that use multiple media and high-speed networks at multiple locations. The idea would be to encourage creative projects that use the high-speed network and HPC facilities. In some cases these projects have brought artists and scientists together to creatively remix the knowledge used by traditional HPC users. It is hard to say what such projects will be, but there is increasing activity in fields interested in experimenting with new media. Such projects often partner with university galleries and museums and offer a way to reach out to the larger community.

### 3.3 Training and outreach

The key to continuing the momentum of the Mind the Gap workshop will be to provide scholars with the training and information they need in order to utilize the resources available. First of all, groups of Humanists who are already involved with HPC projects need to act as evangelists and enablers across the discipline. In practice, this means that there need to be more people who understand the research environment and the technology involved to a reasonable enough level that they can help scholars translate their research questions into HPC appropriate projects. More staff working with DH scholars need exposure to the HPC technologies available at the institutional, regional, and national level so that they are able to effectively guide researchers towards the right technology. A part of this would be to develop high quality training materials, perhaps through the SSHRC ITST programme, working with WestGrid or SHARCnet. This might be in the form of a workshop that develops training materials and test them on user groups with the intention of sharing them amongst the wider DH community. Also, more Mind the Gap-type workshops in which HPC problems are tackled in a collaborative fashion so that CS and DH scholars can see how the process works.

**R.9 Recommendation:** Digital humanists should take the lead in developing **Training** packages for their local facilities. **Common training materials** could be developed by the digital humanities community once for adaptation by others.

The typical HPC training workshop is aimed at an audience in science or engineering, which has traditionally been the audience. We recommend that specialized training be developed that can be delivered by staff familiar with Humanities research. To be honest, good outreach is probably what most consortia need so that humanists learn why they might want to go for training. The Digital Humanities community should lead the development of appropriate training working with their HPC facilities.

There is also an opportunity to develop training materials that can be used across the country for the digital Humanities community. These materials could be developed with a grant from SSHRC and collaboration with an HPC Consortium. They should be released under a Creative Commons license that allows them to be adapted by others. SHARCnet has a good Introductory

essay with examples called "High-Performance Computing in the Arts and Humanities."

**R.10 Recommendation:** The **Mind the Gap model** should be adapted so that it can be offered at other facilities to enable teams across the country to take advantage of Compute Canada.

Further, we believe that Mind the Gap (un)workshops work well to move teams that are ready to adapt to HPC to the point where they have viable projects and can apply for funding. See the section below on running one yourself.

**R.11 Recommendation:** HPC facilities should have **staff familiar with Humanities research** practices who can provide outreach, advise and support researchers new to HPC. Digital Humanists can serve as **mentors**.

HPC facilities need to cultivate enablers or evangelists able to reach out the Humanities and then to help them take advantage of the resources. In other fields this needed less as there is tacit knowledge in the field of what HPC is, what you can expect from such facilities, and how to take advantage of them. Most arts and Humanities researchers don't even know that HPC facilities are available, let alone what they could use them for. Therefore, staff at HPC consortia should be given the opportunity to learn about Humanities research so that they can provide appropriate support. The Digital Humanities community should work with their local facilities to educate support staff and be prepared to act as mentors.

### 3.4 Management

One of the best ways to involve humanists is to not treat them as a community to be served after the design of the facilities. We recommend humanists be involved from the beginning in both the design of facilities and the management of the facilities. We recommend this at both the consortial level and at the national level. While in theory it should be possible for anyone to get voted on to the right committee, the reality of most academic organizations is that people need to be encouraged to come to the meetings and to take on management roles.

**R.12 Recommendation:** Humanists should get involved appropriately in the **development and management** of HPC facilities.

Compute Canada could work with one or more of the organizations that represents the Humanities to ensure that there is useful input into the long term planning around HPC. Obvious organizations that CC could work with are the Society for Digital Humanities (SDH/SEMI), SSHRC, or the Federation of the Humanities and Social Sciences.

### 3.5 Partnerships for Research Support

It seems clear that at this moment humanists need support beyond the ordinary in order to adapt their research to HPC facilities. Such support need not come from Compute Canada. It could come from partnerships with other units like:

- Libraries (for acquisition and long-term digital archiving infrastructure)
- Research Institutes (like the Canadian Institute for Research Computing in the Arts at the University of Alberta)
- Faculties of Humanities or Arts
- Social Science and Humanities Research Council (SSHRC)

**R. 13 Recommendation:** SSHRC and other funding bodies should develop **competitive programs for supporting** HPC projects in the humanities and social sciences.

The NEH has experimented with a competitive grant for compute cycles. We recommend that SSHRC collaborate with Compute Canada to develop a program for humanists (and social scientists) to apply for the research support they need jointly from their local HPC consortium and SSHRC. Were SSHRC to, in some way, create a clear pathway for the SSH community to get funding for HPC research projects that would send a message to both Compute Canada and to CFI that this is an important and innovative direction for research in the arts and Humanities. We recognize that SSHRC and other funding bodies (like individual university research boards) are unwilling to create specialized programs, an alternative would be to flag existing programs as suitable for HPC support. It should be noted that NSERC does provide support to Compute Canada, why shouldn't other research councils?

**R. 14 Recommendation:** HPC facilities should work with their participating universities to develop competitive **Faculty Fellowships** and **Graduate Fellowships** that provide the support needed to develop a project.

SHARCnet has pioneered both graduate and faculty fellowships that provide specialized support to the winning candidates to help them take advantage of HPC facilities for appropriate projects. Such fellowships can be organized in collaboration with other organizations like Humanities research institutes or faculties in order to secure teaching release from the units the fellows come from. Such fellowships do not have to come with monetary awards - they can be a way to award limited programming support to get projects to the point of viability. The important thing about fellowships is that these are an established way to free up people in the Humanities to focus on new and risky research, especially in computing. The award of the fellowship itself rewards the successful candidate for the work of adapting their research to HPC. For graduate students a fellowship is even more valuable as it could provide unique training, differentiation from other graduate students, and encouragement to develop a research agenda enhanced by HPC that could change their career.

**R.15 Recommendation:** The Humanities community have to **make the research and business case** for HPC support. It is our responsibility to show the need and to work with others to develop facilities.

It is incumbent on the Digital Humanities community to make the business case for HPC support for research in the Humanities. It is our responsibility to reach out to facilities and help bridge the gap.

## 4. Next Steps

What are the next steps? Here are some short and long term activities worth pursuing:

- **Examine international models** - How are Humanities disciplines utilising Grid computing and HPC in other countries. How are other Grid consortia engaging with the Humanities community, are they doing it all, and how are Humanists across the world taking advantage of these technologies.
- **Articulate needs of humanists** - Humanists need to fully articulate to ourselves and to the Computing Scientists what our specific needs are with regards to HPC. The list of recommendations about is a good place to start. In addition to a list of needs from HPC, we need to articulate how our need for people to build and populate databases as a platform for Scholarship is regarded as 'infrastructure' in the Humanities as much as microscopes for the Sciences. Though this argument has been framed in the past, and lead to successful CFI grants, we need to continue to refine this argument for HPC specifically.
- **Estimate solutions so HPC folk can see consequences** - As a way of speaking the right language to WestGrid and Compute Canada, we need to translate our needs into specific hardware solutions that can be put in the context of grid computing.
- **Make the business case** - Articulating our needs within the language of Compute Canada is part of the wider need for presenting a good business case for including the needs of Humanists into the NFP-2 application to refresh all Compute Canada nodes. It is crucial that the Humanities take advantage of the opportunity to ensure our needs are part of the long-term Compute Canada strategy. Part of building this case will be more successful Mind the Gap workshops in which tangible results derived from HPC can be
- **Share recommendations and needs with consortia and CC** - The first step in entrenching the Humanities into CC long-term goals will be to share the recommendations of Mind the Gap with all national consortia and the Compute Canada executive boards.
- **Consult with SDH, Federation, SSHRC and other organizations that represent the Humanities** - The best way to get our recommendations shared, heard, and legitimised is through consultation with national funding boards representing the Humanities. It is crucial for our case to be presented to CC and the consortia by these bodies. We should strive to use the success of Mind the Gap to present to SSH organizations the need for personnel support that will result in additional support and pull with Compute Canada.

## 5. How to Run One a Mind the Gap

The following are reflections on what it takes to organize such a workshop.

- **Appropriate Teams.** Mind the Gap was developed for teams that were ready to prototype HPC research applications. All the teams had a programmer and many had computer scientists in addition to Humanities researchers.



- **Proposing Project.** Teams can be asked to apply for the workshop with a research statement describing the project they wish to prototype.
- **Support On-site.** Such a workshop only works if there is excellent on-site support for the duration of the workshop. HPC staff familiar with the facilities and programming for clusters should be on hand to help the teams. Teams will typically need the most help on the first day and it tapers off.
- **Basic Training (but no more).** An unconference/unworkshop should not be structured around lots of training. Most training prepared by HPC facilities is inappropriate for research teams with different skill levels. Instead, in the spirit of an unconference, the organizers can elicit special training or activities that participants want. Given break-out rooms and lots of team time, the special activities can be organized by and for participants.
- **Team Time.** Provide generous slots of time for the teams to work on their prototype - at least half the time. There should be at least 4 hours a day of team time.
- **Break-Out Rooms.** Something we didn't anticipate was the need for break-out rooms for groups to show things to each other and to have more lively discussions. We were lucky that there such rooms available at the Telus Centre.
- **Invited Talks.** Mind the Gap also included invited talks that were open to the broader community. These talks were designed to be of interest to the participants and to take advantage of the expertise gathered. All of the talks were advertised widely and gave an opportunity for those not ready to put together a team to learn about HPC.
- **Catered Lunches and Breaks.** There were long breaks scheduled, but given how teams were working together few people paid attention. If you have catered breaks and lunches many will break when they want to.
- **Sharing Times.** At the end of the day on Thursday the teams had to present their results. This was the capstone event as the Friday morning was set aside for discussion of next steps and recommendations. One suggestion that came from participants was that there should be more structured sharing between teams, starting with short presentations at the beginning of the workshop. There was clearly a lot of informal sharing, including the running of algorithms from one team on data from another, but there could have been regular sharing sessions.
- **Summary Discussion.** MtG included a final morning on Friday dedicated to learning about Compute Canada and discussing next steps for HPC and the Humanities. This session may be needed in other contexts or might be better placed earlier when people are not exhausted as we were.

## 6. Links

- [Web Site](#) for the Workshop.
- For a summary of what High-Performance Computing is and how it can be used by humanists see [High-Performance Computing in the Arts and Humanities](#)
- [Digital Humanities and High-Performance Computing](#) is a wiki with information about the SHARCnet 2008 workshop.
- Geoffrey Rockwell's [Conference Report](#) online.

- [Express News](#) story on University of Alberta site.
- Photos on Flickr <http://www.flickr.com/photos/geoffreyrockwell/sets/72157623924208575/>

---

May 28, 2010