SHARCNET™

SHARC BYTES
newsletter

Winter 2010 VOL. 7 ISSUE 2

# Literary History with a Difference

BY ARMAND VLADAU, STUDENT COMMUNICATIONS OFFICER, SHARCNET

Susan Brown
University of Guelph

Traditionally, high-performance computing has been used primarily in the academic disciplines of physics, astronomy, chemistry, and computer science – fields requiring complex simulations from very large datasets. Recently however, humanities researchers have also started using new computing technologies and, as a result, adding new and unexplored dimensions to the nature and scope of humanities work. Specifically, in 2006 women's literature scholars were presented with a new and exciting computing tool – a type of database capable of allowing researchers to investigate complex research questions. It was in 2006 that *Orlando: Women's Writing in the British Isles from the Beginnings to the Present* debuted and changed the relationship between computing and the humanities. *Orlando* is a unique 'textbase' containing articles on the history, literature, and culture relevant on women writers in Britain. The semantic encoding of its content, allows users to track such matters as the reception of an author's writings, related intertexuality and influence, literary responses, textual features, and other aspects of the author's literary career and life, not just in the entries directly on her, but throughout the textbase. Susan Brown, a co-founder of *Orlando*, is now collaborating with SHARCNET to develop software that will generate three-dimensional information networks based around search results. Brown and SHARCNET will push the *Orlando*'s capabilities even further by enhancing the standard search-and-retrieval text browser with a visual modelling tool for illustrating relationships, which scholars will use to explore new and unexpected paths and patterns.

Brown is a professor of English at the University of Guelph and the current director of The *Orlando* Project. Her original research interests in Victorian writing and its relationship to such matters as feminism, imperialism, and economics have been joined by increasing involvement in the growing field of digital humanities as a result of her involvement in this trailblazing project.

Appropriately named after the Virginia Woolf character, *Orlando* covers a large time period and has changed and expanded since its inception. Not surprisingly, developing such a complex and groundbreaking textbase from the ground up was a long and difficult process. Creating *Orlando* required well over a decade of hard work, primarily because the project was breaking new ground in applying custom semantic markup to extensive, collaboratively authored, digital content, and because its delivery system was developed to work seamlessly with the markup. The final product, peer-reviewed and published online in 2006, is a valuable electronic resource focused on the lives and writings of female writers in Britain over the past three hundred years. It is an impressive achievement for humanities scholarship and a tremendous step forward for the new field of digital humanities.

*Orlando* has successfully transitioned from the restricting medium of print to the easily and widely accessible digital medium. Its predecessor, *The Feminist Companion*

Continued on page 6

## in this issue:

# Message from the Chair of the Board

Paul Maxim,
Chair of the Board

"Simulation supported by high performance computing infrastructures has become the third pillar of science, complementary to experimentation and modeling."

The G8 Research Councils Initiative on Multilateral Research Funding

It is my pleasure to present another issue of SHARC Bytes to the broader research community, institutional representatives and our funding agencies.

Over the years, SHARCNET has become a multifaceted high-performance computing (HPC) organization enabling a widening range of research needing HPC for innovation and competitive advantage. The number of researchers and projects that rely on SHARCNET as a critical resource continues to grow, particularly in the non-traditional HPC disciplines.

SHARCNET resources include: leading computational hardware, software and visualization capabilities; programmes for new and advanced users; skilled technical personnel; virtual collaborative research spaces; research fellowships and excellence-based resource allocations; faculty chairs taking HPC to the next level; and outreach that brings the significance of advanced computing to the public, high-schools and undergraduates. SHARCNET also provides a powerful platform for interactions with industry. We support many academic researchers collaborating with industry and, increasingly, provide expertise directly to companies and government organizations.

SHARCNET, together with its sister organizations HPCVL and SciNet provide an advanced HPC platform to support forefront research in Ontario. The three HPC consortia span the province and include all research universities, several hospitals, colleges and research institutes. Together with the provincial optical research and innovation network, ORION, which links all educational institutions and infrastructure, the networks comprise a powerful provincial platform for advanced research using HPC that links with peers in Canada and internationally.

SHARCNET has, as one of the largest HPC consortia in Canada, taken a leadership role in advancing the objectives of the Long-Range Plan and HPC generally across Canada, as part of the Compute Canada initiative. Compute Canada is the national organization that coordinates and promotes the use of high performance computing in Canadian research. In collaboration with Canada's seven university-based HPC consortia (WestGrid, SHARCNET, SciNet, HPCVL, RQCHP, Clumeq and ACEnet), it works to ensure that Canadian researchers have the computational facilities and expert services necessary to advance knowledge and innovation across disciplines. Compute Canada is currently waiting for the next National Platforms Fund (NPF) call for proposals, which is anticipated later in 2010. This will ensure a much needed boost in Canada's HPC infrastructure.

SHARCNET is now a mature HPC organization that provides high quality service to a broad base of researchers – ranking as one of the top HPC Centres in Canada. However, to stay in this position, it is important that funding for HPC continues, both at the federal and provincial levels. We believe that the use of HPC is far from saturation, and in this context, we expect user demands to continue to increase in future. It is, therefore, imperative that we continue to work with Compute Canada and the Ontario HPC Council to ensure that these essential resources are sustained at a competitive level.

Paul Maxim
Chair, SHARCNET Board of Directors
Associate Vice-President: Research
Wilfrid Laurier University

# Scientific Director's Message

Recent months have seen SHARCNET investing a great deal of time and effort in providing increased support for and leadership of national and provincial HPC initiatives. An Ontario High Performance Computing Council was formed in recent years to provide effective oversight of the province's academic HPC capacity. Assisting in this effort is the HPC Coordinating Committee, consisting of the directors of the three provincial consortia, SHARCNET, SciNet and HPCVL. As you are probably aware, both SHARCNET and HPCVL applied for renewed funding through the Ontario Research Fund - Research Excellence programme last July. We have recently received word that we will not receive funding through this programme. We have funding to support operations well into next year and will endeavour to operate our infrastructure for as long as we can. I remain optimistic about the continuation of SHARCNET and new funding but this will require significant effort and cannot be taken for granted. I cannot emphasize too strongly the value of community involvement in stressing to funding agencies and senior administrators the importance of HPC for a competitive research environment. I urge you to become involved in this advocacy: the next few months will be critical in this regard. In line with raising awareness of the often tenuous nature of HPC in Canada, this issue of the newsletter has an article examining Canada's international HPC position. I will update you on progress over the months ahead.

The development of Compute Canada is proceeding apace. One of the key efforts has been to establish a Canada-wide user accounts system. This will enable users to more easily access any of the Compute Canada resources available to them as part of the National HPC Platform. We are therefore requesting that all existing SHARCNET users, who have not already done so, create a Compute Canada account and link it to their SHARCNET account. To register for a Compute Canada account, visit the Compute Canada Database server at: ccdb.computecanada.org/security/login

We are delighted to be a sponsor of the ORION Summit, Ontario's premier research and education conference at the MaRS Centre in Toronto, April 12-13, 2010. A number of Ontario's top research and education leaders are converging at the annual event to meet and network with a unique line-up of innovators in science, business and advanced education, as well as champions of new and interactive and collaboration technologies and more. SHARCNET and Compute Canada have partnered to be gold sponsors of the Summit, and will host the talk: "High Performance Computing – Who's using it, who should be and should you?" We hope to see you there.

In November, 2009, the Dedicated Resources Programme, which allows researchers to apply for dedicated CPU time and storage, was revised to encompass two streams with different timelines and thresholds, now referred to as "Small DR" and "Large DR".

SHARCNET recently issued a call for proposals for Round XII of SHARCNET's Research Support Programmes: Large DR, with a deadline of April 30, 2010. Small DR has a competition every two months; the next application deadline is April 15, 2010. For more information, please visit www.sharcnet.ca/my/research/dedicated

On the hardware front, I am pleased to provide an update on the following new systems/services:

- The new Grid Lab, purchased from SHARCNET CFI Round 4 funds and hosted at Windsor, was completed in February.
- The RFP for a new cluster, to be purchased from SHARCNET's allocation from the CFI National Platform Fund, closed at the end of February. This will result in a new cluster about the same size as "Saw" but with newer processors and should be available to users early this Summer.
- The new "global work" storage hardware has been delivered and should be available to users by the summer.
- The second AccessGrid rooms at Guelph and Western are now operational. These rooms are much larger and, in the case of the Guelph room, will be used to support course offerings such as the parallel programming graduate course.
- New visualization workstations have been installed, which will drive a 3D projector system at Lakehead, a 3D display at McMaster and a large tiled display at Guelph as well as labs at both Western and York.

Plans are also underway for "SHARCFest 2010", our annual outreach campaign of open houses, training events and symposia starting in May to help educate the broader community on the uses and importance of high-performance computing to research. We are currently working on a number of activities including SHARCNET Research Day 2010, planned for May 6th at York University. The anchor event for SHARCFest will be our annual HPC Summer School scheduled from May 31-June 4 at Sheridan College. Watch the SHARCNET website and events calendar for more information!

Hugh Couchman,
Scientific Director

Hugh M.P. Couchman
SHARCNET Scientific Director
Fellow, Canadian Institute for Advanced Research
Professor, Physics and Astronomy, McMaster University

# Computer Technology Helps Break Ground in Cancer Prevention

BY REBECCA HANNAM, STUDENTS PROMOTING AWARENESS OF RESEARCH KNOWLEDGE (SPARK) PROGRAM AT THE UNIVERSITY OF GUELPH

Wely Floriano, Lakehead University

Helping doctors assess if some patients have a higher risk of developing cervical cancer is the goal of a new research project at Lakehead University.

Dr. Wely Floriano, a member of the Biorefining Research Initiative and Thunder Bay Regional Research Initiative, is working through SHARCNET to use computational tools to study proteins related to the human papilloma virus (HPV). This virus infects the skin and genital areas, and can eventually lead to cervical cancer in some women.

The ultimate goal of Floriano's work is to discover an optical imaging probe that will allow doctors to detect cervical cancer before the disease fully develops. The probe is a fluorescent compound that could detect a protein produced by HPV, which acts as a biomarker for cervical cancer. Imaging the cervix after a local application of the probe with, for example, a low-light or high resolution digital camera attached to a colposcopy, will allow for the identification of tissue at high risk of developing cancer.

"Using the probe to identify high-risk areas prior to development of the cancer will be important for improving the prognosis for treatment, and will allow doctors to preventatively treat patients," says Floriano.

Her first step is to find a fluorescent chemical compound that can detect certain variants of the protein E6, a protein produced by HPV which could indicate the probability of cervical cancer developing. To this end, Floriano is using a computer program to screen a large database of chemical compounds against the target protein. The process uses computer-generated three-dimensional structures for the chemical compounds and an experimentally determined structure of the protein.

The computational tools used in this process simulate the interactions of each chemical compound with the HPV-related protein. The computer is able to estimate the energy of these interactions, and determine the chemical compounds that interact most effectively with the E6 protein. Compounds that are successful in this process are then tested experimentally to confirm target protein binding. Using a combination of computational and lab work, an experimentally confirmed "hit" can then be developed into an imaging probe.

An optical imaging probe that could detect cervical cancer would be a major advancement in preventing HPV-related cervical cancer. Doctors would be able to use the probe to assess the probability of cancer developing in patients diagnosed with HPV.

Because various strains of HPV exist – and some indicate a higher risk of cancer development than others – tissue from a patient's cervix can be screened to determine which types of HPV proteins are present. If this assessment shows a patient has a high risk type of HPV, doctors can use the imaging probe to map the cervix and identify regions at risk.

In turn, they will be able to provide the patient with preventative treatment specific to cervical cancer.

Floriano's research projects uniquely focus on developing and applying computational tools in oncology. Her professional team uses many sophisticated computer programs to learn more about chemical compounds and how they can be applied to advance modern medicine.

"Starting research projects with a computational approach is fast and efficient," says Floriano. "Initial lab work can be time consuming, but with computer technology, only the most promising compounds need to be tested experimentally."

Floriano credits SHARCNET for making many of her biomedical projects possible. SHARCNET provides researchers with access to sophisticated computational resources typical to large pharmaceutical companies, so she and others are able to tackle major medical projects significant to society, such as making a non-invasive risk assessment test for cervical cancer in HPV patients a reality.

Funding for this project is currently provided by the Thunder Bay Regional Research Initiative. Dr. Floriano is also a SHARCNET Research Chair, with funding provided by SHARCNET's vendor partners, HP and SGI.

# HPC: Who Needs It and How Much is Enough?

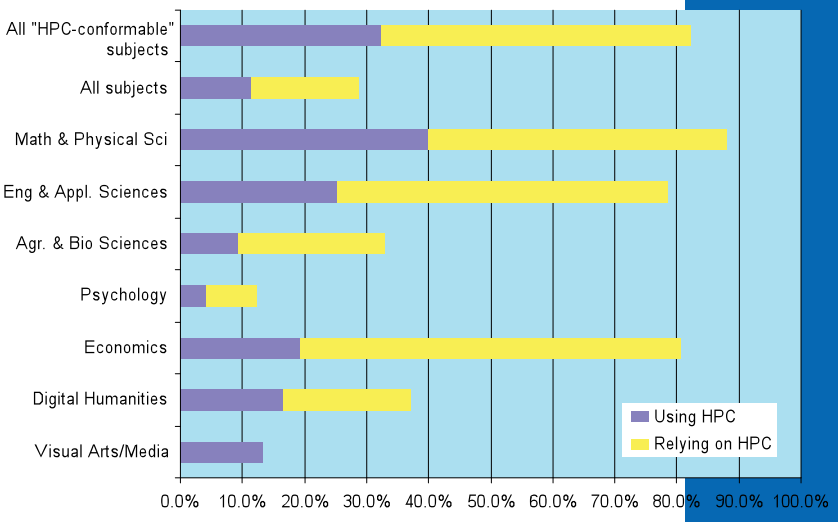
In a recent special report, the Economist highlighted the challenges posed by the vast amounts of data that are now being collected in areas as diverse as banking and finance, the environment, national security, retailing, subatomic physics and astronomy[1]. Among the challenges, however, the report made almost no mention of the substantial computer power that is needed to manage and analyse these data. Similarly, a Google search, with apparently little effort, can generate millions of hits in a fraction of a second without the user's knowledge of Google's massive back-end server farms. For datasets of a few Terabytes or more, very modest by today's standards, substantial computational muscle is needed for any intensive analysis; in many cases high-performance computing (HPC) is essential. These examples illustrate an important difficulty facing providers of information and communications technology (ICT - from networks to HPC): it is largely invisible. Understandably, users focus on the end goal; the enabling technology is taken for granted until it fails or underperforms.

> "HPC as an enabling technology is taken for granted until it fails or underperforms."

This article examines the challenge of providing HPC in the context of shared resources at Compute Canada and SHARCNET. Given the uptake of HPC by a wide range of disciplines internationally, we ask how ubiquitous is the need for HPC in Canadian research and is the technology provisioned at a level that will allow researchers to compete internationally? Specific needs for HPC can be identified fairly readily in various disciplines, genomics, for example, is critically dependent on HPC; what is less clear is the aggregated need for HPC across the whole of the Life Sciences. Similarly, what is the aggregated need for HPC across all disciplines for the whole research community in Canada? We address this in two ways: first, how widespread is the use of HPC by various disciplines and what is the overall impact in those disciplines? Second, how does the provision of HPC in this country compare with that of our G8 and G10 competitors? Note that the precise flavour of the HPC needed, traditional supercomputers, grids or clouds, is a detail that is of secondary importance: the underlying hardware is essentially the same whatever modality is used and the need for effective systems- and user-support remains unchanged.

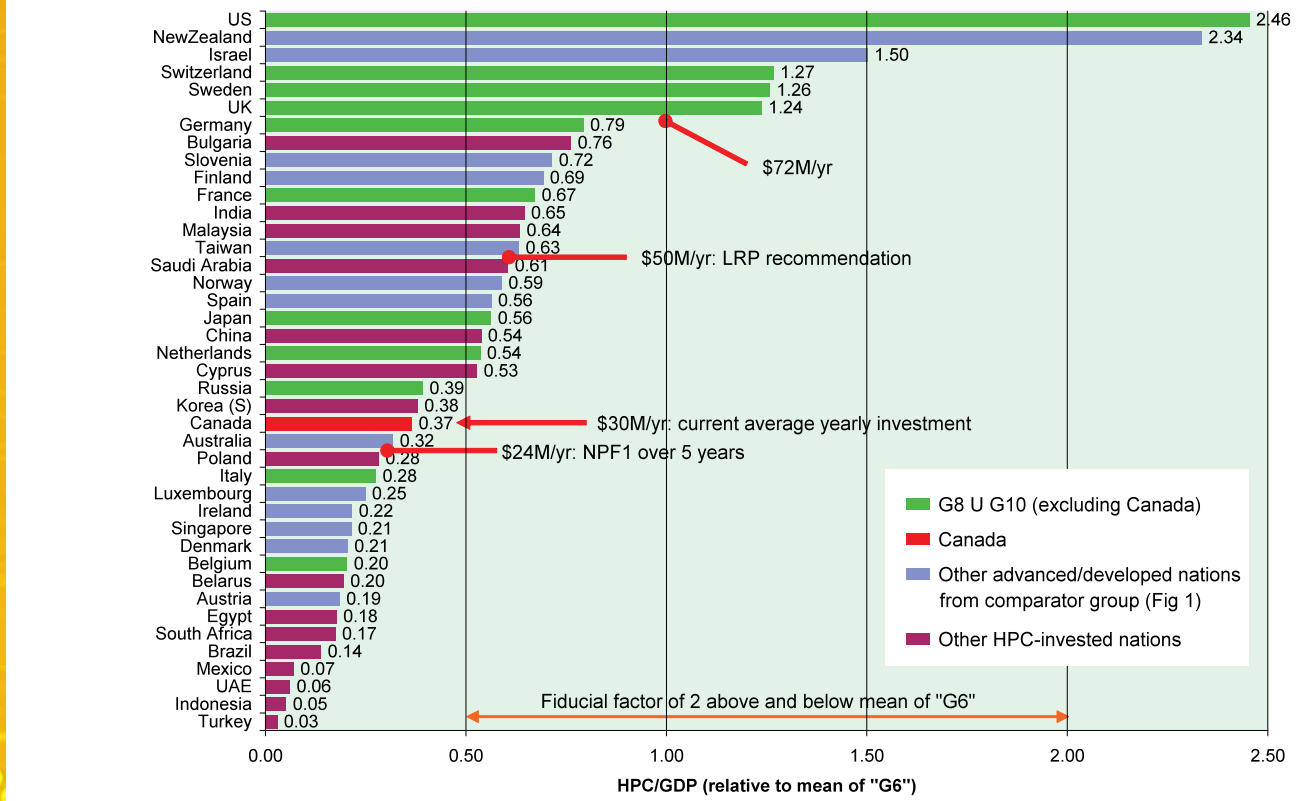
The first figure shows, by discipline, the percentage of all faculty at the SHARCNET member institutions who have SHARCNET accounts (purple bars). For faculty in the Humanities and Visual Arts, the (small) fraction who are likely to benefit from HPC has been estimated. The story does not end there, of course: in the same way that theorists and experimentalists frequently collaborate, and yet only the latter require lab space, so computational researchers are an increasingly necessary third estate for successful research teams. This effect was estimated by surveying SHARCNET users for information about colleagues who have benefitted from SHARCNET through such collaborations (yellow bars). Overall, a significant fraction of researchers now depend on HPC[2].

**Fraction of SHARCNET faculty using and relying on HPC**




We also looked at the availability of HPC in Canada and compared this with the corresponding investments in other countries. The second figure shows a rank ordering of countries according to the installed HPC base normalized by GDP. Normalizing a wide range of research measures by GDP - including research funding, citations, papers (and population) - leads to remarkably uniform level of activity and output for nearly all G8 and G10 countries, certainly within a factor or two above and below the mean of the six largest economies. The figure shows the challenge faced by computational research in Canada. Although investments by CFI and the provinces dramatically changed the landscape for HPC in the early 1990s, initiating a sustained investment of approximately $30M/year, it is still significantly below the mean. With the current timetable for the second round of the National Platforms Fund, the rate of investment has fallen to roughly $24M/year, a third the level of our G8 partners.

Two messages are clear from these studies: HPC is both fundamental to a very significant fraction of the research community in Canada and is underprovided relative to many of the developed countries that are our intellectual and economic competitors. The relative lack of investment in HPC is mirrored by the more general under-investment in ICT by business that was




recently identified by the Council of Canadian Academies[3] as a major reason for Canada's productivity gap. Although HPC is a very specific component of the ICT landscape, it is the critical vehicle for creating knowledge from the vast amount of data being collected or generated. In an era of sophisticated modelling of complex systems and the necessity of analysing massive datasets, it is critically important to accord HPC a visibility that is commensurate with its importance to a modern knowledge-based economy. If modern optical networks are the 21st century equivalents of roads and railways, then HPC corresponds to the major cities that generate and use the products that flow on these transportation arteries... and that produce the wealth of nations.

1. Economist Feb 27th-Mar 5th, 2010
2. .Details of the SHARCNET studies may be found at: www.sharcnet.ca/my/documents/index/General
3. "Innovation and Business Strategy: Why Canada Falls Short", April 2009 (www.scienceadvice.ca/innovation.html)

Literary History with a Difference, continued from page 1

*to Literature in English*, edited by Virginia Blain, Patricia Clements, and Isobel Grundy, was published in 1990 in print form, but suffered from a number of restrictions. The print companion, a dense 1,231 double-column pages, was bursting with content and afforded only a minimal index. The transfer to the digital medium provided a number of benefits either not possible or not adequately provided by print. *Orlando* is not constrained by the capacity of a binding, it can be expanded and updated as needed, and web delivery makes the material accessible from anywhere in the world. Over the last four years, *Orlando* has gained a reputation as a reliable, dynamic, robust and innovative scholarly resource, allowing users to search for information and critical interpretation ranging from an author's name or pseudonyms, to the settings, motifs or characteristics of little-known texts, to the social and literary climates of the time. The textbase is so thorough that a user can, for example, "trace Virginia Woolf... through all the entries on other writers that mention or discuss her, grouping those mentions by context: whether she appears there as another's friend or associate, or as publisher, reviewer, literary influence, or whatever". If it was physically printed today, the still-expanding *Orlando* would easily fill over seventy printed volumes with the current content alone.

But it is not so much *Orlando*'s size as its computationally accessible structure that makes it a unique testbed for developing tools to exploit effectively large quantities of scholarly information.

Continued on page 7

Literary History with a Difference, continued from page 6

"Working electronically", the team explains in the article, "creates a degree of cross-referencing and textual inter-relation impossible with print scholarship. *Orlando*'s semantic markup also makes its materials amenable to inquiry with computers in a whole range of ways we have yet to explore fully". A recent review in *Eighteenth-Century Fiction* commented that "*Orlando* encourages the researcher to see new patterns, new connections, and new traditions–and thus to think in new ways." The new visualization tool will soon further enhance the already innovative and exciting user navigation experience by literally transforming how users see the materials.

Although *Orlando* has been live and in use since 2006, only a handful of beta users have so far tested the new visualization interface being developed by Brown and SHARCNET. In the paper *Visualization for Literary Historical Analysis*, it is stated that the goal of the project is to "assist literary scholars working with *Orlando* materials by designing systems to support speculative inquiry, by leveraging text markup, text mining, and visualization". The collaboration with SHARCNET will generate a more intuitive navigation system, shifting from the traditional text search-and-retrieval model to a visualization model using three-dimensional graphs.

Brown has been collaborating with Mike Bauer, Department of Computer Science at Western and SHARCNET Associate Director, and Jennifer Berberich, a recent Computer Science graduate, to generate visualized search results. The team is using tags and nodes to create software that will clarify certain patterns or connections between search terms, allowing users to explore the lives and writings of women writers in the British Isles in new ways.

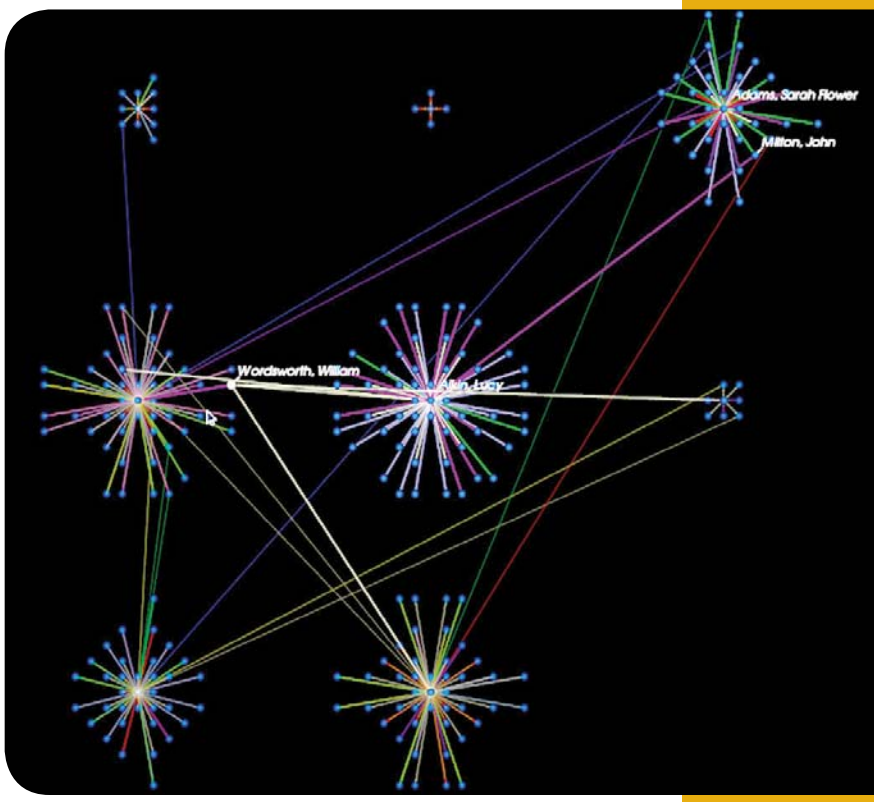
In *Visualization for Literary Historical Analysis*, the team provides a complete breakdown of the new features of the navigation system that will be added to *Orlando*. The interface provides four different navigation modes, Graph Mode, Toggle Mode, Highlight Mode and Camera Mode, which will allow users either to broaden or focus the scope of their results by selecting which information to include and which to exclude or 'dim'. Perhaps most visually exciting is the Camera Mode in which a user can view and rotate a graph in three-dimensions, giving a user the ability to move around and explore every connection.

**A simple figure depicting the network of nodes (individuals) and edges (relationships) generated by a search query using *Orlando*.**

Ultimately, the aim of the collaboration is to export the visualization software developed at SHARCNET to a web-based delivery system. This will give users access to the textbase visualization interface from any personal computer, minimizing geographical and technological barriers to accessing *Orlando*.

As a testament to what can be achieved by humanities computing, a potential application of *Orlando* is that it could be used as a blueprint for developing future textbases with different academic focuses. Indeed, the *Orlando* systems will form the foundation of the Canadian Writing Research Collaboratory, a CFI-funded platform to enable literary scholars to collaborate in the production, analysis, and dissemination of digital research materials. Brown and her *Orlando* colleagues have created a benchmark in digital humanities research that will greatly benefit literary scholars and future academic scholarship on women's literature. The immediate application of this new work is that it will give academics access to the complete *Orlando* textbase visualized from any home computer. Users will be able to focus, refine, and explore intricate and important networks relating to women's writing in the British Isles from the middle ages to the present.

About *Orlando: Women's Writing in the British Isles from the Beginnings to the Present, co-edited by Susan Brown, Patricia Clements, and Isobel Grundy, and the result of the collaborative interdisciplinary efforts of more than 100 researchers* is published by subscription on the web by Cambridge University Press. For more information on Susan Brown or the *Orlando* project, please visit: orlando.cambridge.org and www.ualberta.ca/ORLANDO/



# High Performance Computing Symposium 2010

SHARCNET is pleased to highlight HPCS 2010, being held June 6-9, in Toronto, and hosted by SciNet. HPCS (High Performance Computing Symposium) is Canada's foremost HPC conference - a multidisciplinary conference where computational researchers from all disciplines in industry and academia, computer scientists, and vendors exchange new tools, techniques and interesting results in and for HPC computational research. The HPCS 2010 conference will take place at the University of Toronto on June 6-9, with workshops on the two days preceding the symposium.

Please visit www.hpcs2010.org for more information.

# SHARCNET Launches New Small DR Programme

The Dedicated Resources Programme has been revised and it now encompasses two streams with different deadlines and resource thresholds. The "Small DR" stream is new and will be run as a pilot initially to assess its utility. The previous Dedicated Resources Programme becomes "Large DR". The Dedicated Resources Programme was established to enable and facilitate computational projects of exceptional research merit that will achieve discoveries of international significance through the optimal exploitation of SHARCNET's computing infrastructure. The programme allocates dedicated computational resources such as CPU time and storage.

Small DR and Large DR have their own deadlines and resource thresholds as depicted in the following table:

| Resource | Resource Threshold | |
|---|---|---|
| | Small DR | Large DR |
| Dedicated CPU Time | 1% – 5% of system | > 5% of system |
| Dedicated storage | More than 15TB for more than 6 months | |
| Deadlines | 15th Feb, Apr, Jun, Aug, Oct, Dec | September, March |
| Allocation window | 3 months | 6 months |
| Project attributes | Urgent, high-priority | Substantial, grand-challenge |

Applications for dedicated resources are open to all faculty (PI) applicants with a valid SHARCNET account.

For more information on the programme or to see a listing of previous awardees, please visit: www.sharcnet.ca/my/research/dedicated


# SHARCNET Research Day 2010

Please join us for SHARCNET Research Day 2010, to be held on May 6th, at York University. SHARCNET Research Day is the premier annual event at which SHARCNET professors, postdocs and graduate students meet to learn about each other's HPC-related research. The theme of the meeting is "HPC Innovation for Research."

We invite you and your research group to attend and present your HPC-related work. Pease visit the Research Day website to register for the event and to submit abstracts for oral presentations and posters. The event is free and lunch will be provided.

SHARCNET Research Day 2010 will feature two keynote presentations by leading HPC researchers, namely:

- **David H. Bailey, Lawrence Berkeley National Laboratory**, Computing as the Third Mode of Scientific and Mathematical Discovery
- **Mikko Karttunen, The University of Western Ontario**, Modelling Membranes, Proteins and Biology: Studying the fundamentals by computer experiments

There will also be several parallel tracks of contributed presentations, and a poster session with prizes for best posters in several categories. Important Dates:

- April 16: deadline for abstracts – contributed oral presentations
- April 23: notification of acceptance for contributed oral presentations – release of final program
- April 29: deadline for online abstract submission – posters
- April 29: deadline for registration (registration is free but required)
- May 6: SHARCNET Research Day 2010 at York University, Curtis Lecture Hall

Visit www.sharcnet.ca/Events/RDay2010 for more information. We look forward to seeing you there!

# SHARCNET Welcomes New Research Chairs

Wely Floriano

Christopher Collins

With funds provided by our private sector partners, Hewlett Packard and Silicon Graphics Inc., SHARCNET was able to allocate funding for eight new SHARCNET Research Chairs, including one College Research Chair. Several Research Chairs are already in place, and we are pleased to welcome the following new incumbents:

**Dr. Wely Floriano**,
Biorefining Research Initiative, Lakehead, effective August 1, 2009

As is highlighted in this issue of SHARC Bytes, Dr. Floriano's research involves the development and application of computational tools to simulate biologically relevant systems. In the context of Medical Biotechnology, these computational techniques are applied for the development of new and modified chemical entities, such as new medicinal drugs and taste modifiers. In the context of Environmental Biotechnology and as part of the Biorefining Research Initiative at Lakehead University, computer-assisted molecular design methods are used to study genetic sequences of biomass-degrading microorganisms for selection and/or modifications aimed at improving efficiency and reducing waste of biorefining processes. These techniques are also used to identify economically relevant targets for chemical compounds produced as by-products of biorefining processes or isolated from forest resources, and to suggest chemical modifications to primary compounds to make them more valuable (e.g., a primary compound found to be a weak agonist of a pharmacologically relevant target may be further designed into a marketable medicinal drug).

**Dr. Christopher Collins**,
Department of Computer Science, University of Ontario Institute of Technology, effective January 1, 2010

Recently appointed SHARCNET Research Chair, Dr. Collins' research involves visualizing natural language processing. Textual and speech data are at the forefront of information management problems today. From email to case law to call center speech recognition transcripts, managing and filtering the volumes of linguistic content to find subsets of interest and to understand it is an immense undertaking that most of us are faced with everyday, when we open our inbox or use our favourite search engine. Linguistic analysis is also very important to business as the corporate world strives to improve text analytics for monitoring customer opinions, to streamline document management and retrieval, and to provide data services to clients. Drawing on his background in both computational linguistics and information visualization, Dr. Collins' vision is to create solutions to enhance the communication and understanding of linguistic data.

# The Changing Face of New User Training

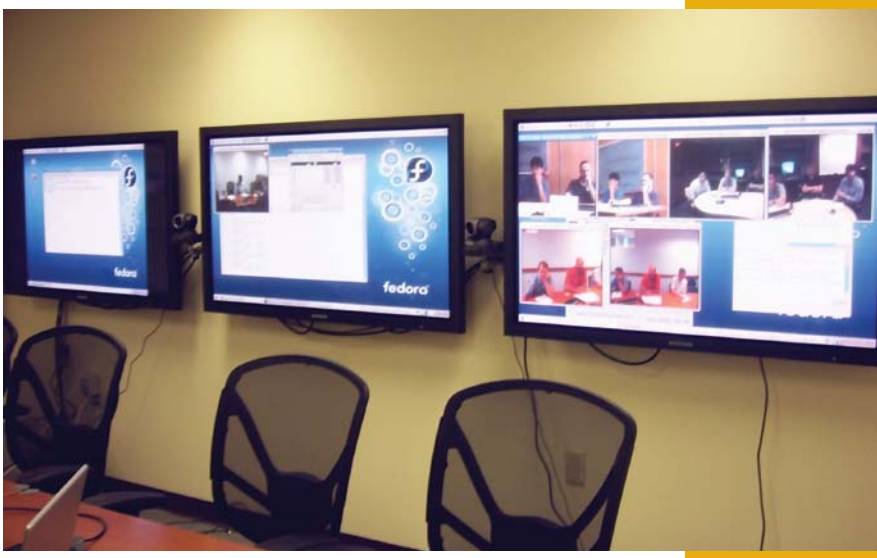
BY DAVID MCCAUGHAN, HPTC CONSULTANT, SHARCNET

When SHARCNET first introduced the user training mandate and started offering regular new user seminars, they were delivered over AccessGrid. This software is wonderful for larger-scale collaboration but can be problematic for users to set up on their desktop computers. In an effort to make the experience as convenient as possible for the user community, we moved these seminars to a webinar format, using software by DimDim. Users could now attend these literacy seminars from any web browser which proved popular with the community.

Some initial technical hiccups, together with feedback from the user community who felt that it would be more convenient for them to just watch a recorded seminar, forgoing the possibility of live interaction for the ability to receive the material on demand, is taking us in a new direction yet again. Starting in April, we will be providing an online recording of the literacy seminar so that new users will be able to receive literacy training on-demand, and eliminating the two-week cycle of new user seminars as it exists currently. New users will be able to apply for an account, and once approved by their sponsor, watch the new user seminar and take their certification quiz at any time, which we hope will further streamline the training process for our user community.

One aspect of the live new user seminars that will be missed is the ability to ask questions. This is important for many users as people are often not just new to SHARCNET, but new to high performance computing itself and can be bursting with questions regarding the use of SHARCNET facilities with their specific software or technology. To better support this need, SHARCNET will be introducing a "virtual help desk" coinciding with the transition away from live new user seminars.

The SHARCNET Virtual Help Desk will take the form of a live two hour meeting, scheduled regularly twice per month, using our DimDim webinar software and staffed by one of our HPC software analysts. Users will be able to use their web browser to connect in and have live face-to-face time with a staff member including audio, video, a shared web browser and the ability to share desktops for demonstration purposes as needed. If there is more than one person online, others can hang around, watch and listen if they are interested while waiting for their turn. Multi-way voice communication is also an option.




The objective of the virtual help desk is to provide a venue for users to ask general questions and get hands-on assistance with basic SHARCNET usage: transferring files, submitting and managing jobs, compiling applications, disk management (including use of archival storage), elementary debugging and any other support facilities we provide. Due to the limited time and online format, this will not be an appropriate venue for detailed help with design or other significant software issues; however, that has traditionally been something for which you would speak directly to an HPTC anyway. The virtual help desk is intended to deal with "frequently asked question" issues for which a user would prefer an interactive venue.

SHARCNET continues to strive to enhance the lines of communication between our users and staff and to provide novel and innovative training opportunities for the researcher community. We hope these new offerings will be well received and further enhance the value of SHARCNET to our user community.

# Swimming with SHARCS: Our HPTC Team

Meet the SHARCNET HPTC Team! SHARCNET's High Performance Technical Consultants are a valuable resource for the researcher community. They are involved in many aspects of supporting HPC beyond the daily user interaction providing code support. These activities include creating an extensive amount of training materials (see www.sharcnet.ca/help), conducting topical seminars as well as general introductory sessions, packaging and maintaining software tools on the systems, and outreach activities to both traditional and non-traditional departments. Questions or issues for an HPTC and you aren't sure who to contact? Email help@sharcnet.ca or submit a problem ticket via the SHARCNET web portal.

"The Technical Consultants are the backbone of the support services provided by SHARCNET."

**Nick Chepurniy** (Windsor)

Applied mathematics, numerical analysis, HPC systems, optimization and performance tuning

Phone: 519-253-3000 x4883
E-mail: nick@sharcnet.ca

**Baolai Ge** (Western)

Applied mathematics, scientific and technical computing, distributed systems

Phone: 519-661-2111 x88544
E-mail: bge@sharcnet.ca

**Weiguang Guan** (McMaster)

Visualization, medical imaging, pattern recognition, image processing/analysis, video processing & analysis, machine vision

Phone: 905-525-9140 x22540
E-mail: guan@sharcnet.ca

**Jemmy Hu** (Waterloo)

Computational chemistry, physics, parallel programming, HPC systems

Phone: 519-888-4567 x37854
E-mail: jemmyhu@sharcnet.ca

**Alexei Razoumov** (UOIT)

Astrophysics, computational fluid dynamics, numerical radiative transfer, adaptive mesh refinement methods, scientific computing

Phone: 905-721-8668 x3853
E-mail: razoumov@sharcnet.ca

**Sergey Mashchenko** (McMaster)

Computational astrophysics and cosmology, parallel computing

Phone: 905-525-9140 x27663
E-mail: syam@sharcnet.ca

**Doug Roberts** (Laurier)

Mechanical engineering, bioinformatics, computer algebra and grid

Phone: 519-884-0710 x3073
E-mail: doug@sharcnet.ca

**Dave McCaughan** (Guelph)

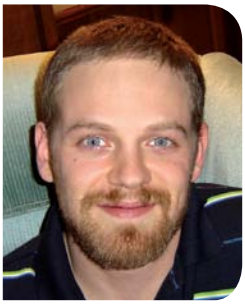
Software engineering, HPC systems, systems programming, computational theory

Phone: 519-824-4120 x56467
E-mail: dbm@sharcnet.ca

**Tyson Whitehead** (Western)

Mathematics, statistics, and computer programming

Phone: 519-661-2111 x82846
E-mail: tyson@sharcnet.ca



**Hugh Merz** (Laurentian)

HPC systems, HPC accelerators, astrophysics

Phone: 705-675-1151 x2347
E-mail: merz@sharcnet.ca

**Darryl Willick** (Lakehead)

HPC systems, computational chemistry

Phone: 807-343-8996
E-mail: dwillic1@sharcnet.ca

**Pawel Pomorski** (Waterloo)

Molecular dynamics methods, density functional theory methods, biophysics, condensed matter physics

Phone: 519-888-4567 x38458
E-mail: ppomorsk@sharcnet.ca

**Isaac Ye** (York)

Computational fluid dynamics, turbulence, and reacting flows

Phone: 416-736-2100 x22935
E-mail: isaac@sharcnet.ca

stereoscopic image using 3D projector

"SHARCNET provides more than just raw compute power as evidenced by the growing list of non-compute tools."

# Non-Compute Tools

In addition to HPC resources, SHARCNET provides a number of tools and services that can be of use to academic researchers. One of them is a Subversion server which provides version control repositories to all our users. You can work with this server either from a login shell on a SHARCNET machine, or directly from your desktop. To access your repository, first you need to check out its content to a local working directory by entering the command "svn checkout https://subversion.sharcnet.ca/svn-username/repo workingDirectoryName" on a SHARCNET machine or directly on your Linux or Mac desktop, where "username" is your SHARCNET username. There are also several Subversion clients available for Windows. Where you put your working copy is entirely up to you, and you can even have several working copies on different machines provided that you bring changes from these copies to the repository with "svn commit". Of course, for a single user all of this functionality can be implemented with private repositories accessed via ssh without a server. The Subversion server becomes really useful when several users work on the same code simultaneously from different computers. If you are a SHARCNET PI, all members of your group have read-write access to your repository at subversion.sharcnet.ca/svn-Plusername/repo. Note that in current implementation only a single repository exists for each user, additional projects can be added simply as subdirectories.

Git is another version control tool being used widely on SHARCNET. Unlike working copies in Subversion, each copy of a code under Git control contains full revision history. To collaborate with Git, two users can set up a clone of their repositories to which they both have read-write access. For users inside a single research group this can be done with file permissions, so that no one outside the group can copy the code.

SHARCNET users have access to several database management systems, such as SQLite, PostgreSQL, and MySQL, which are installed on several of our clusters. These systems include most of the standard features of SQL, along with various extensions. SQLite is a serverless system: a process that wants to access the

database reads and writes directly from the database files on disk, instead of relying on a separate server. Therefore, programs that use SQLite do not need a special setup before they are run. A user can create, start populating and querying a new database either by entering SQL commands manually with the "sqlite3" terminal front-end, or by calling SQLite functions from their favourite programming language, without any administrative support. For small databases SQLite is probably the easiest system to set up and use, and therefore it should be the first option to consider. On the other hand, PostgreSQL and MySQL require separate servers which process all incoming queries. To request a PostgreSQL or MySQL account and a database on one of our clusters, please send us an email at "help@sharcnet.ca" specifying which database system you would like to use, along with some details on the size of the dataset and complexity of the queries.

Scientific visualization is an important part of HPC. In addition to HPC clusters, we have dedicated visualization resources including a rendering GPU cluster at McMaster (rainbow), around two dozen visualization workstations installed at our partner institutions, as well as a stereoscopic projector at Western and a quad-HD screen at McMaster. Our visualization workstations are fast quad-core machines with 8-16 GB of on-board memory and large professional displays - either dual or quad 24" or a single 30" monitor - driven by NVIDIA graphics cards. They are listed under "Specialty Systems" on the SHARCNET website and are suitable for demanding interactive GUI applications such as visualization of large datasets. You can log into these systems with your usual SHARCNET username/password, and they can be also accessed remotely for interactive work via ssh. The stereoscopic projection system installed in the SHARCNET Boardroom at Western allows researchers to study three-dimensional datasets such as magnetohydrodynamical flows in star formation. At McMaster, we have two 22" displays that are capable of 3D visualization using polarization and shutter glasses, respectively, as well as a quad-HD 52" screen. A number of commercial and open-source visualization packages are installed on our systems - for a full list please scroll to the bottom of our software page www.sharcnet.ca/my/software.

SHARCNET users can now have their research website hosted on a web server on "Mako" where they can publish research results, host medium-size datasets for download by the community, put up an interface for job submission via CGI scripts, etc. This website is accessible via username.sharcnet.ca, and we can also host a domain name for a group, e.g., www.researchproject.ca if it owns the domain. Currently website hosting is available to PIs only and can be turned on by individual request sent to "help@sharcnet.ca", although we can consider opening this service to non-PI users on a case-by-case basis. By default these virtual hosts use Drupal content management system (CMS) for easy website administration, and we can add other CMS's and wikis by request. Perl, PHP, and Ruby on Rails are also available. In addition, we can supply empty databases for websites through MySQL or PostgreSQL. We hope these websites will be useful to the community, and if you have any questions or requests, please email us for help, and SHARCNET staff will be happy to assist you.

SHARCNET™

# Shared Hierarchical Academic Research Computing Network

## Board of Directors

Paul Maxim (Chair), Wilfrid Laurier University
Steven Liss (Vice-Chair), University of Guelph
Hugh Couchman (Scientific Director), SHARCNET
Cindy Munro (Secretary), SHARCNET
Ted Hewitt, The University of Western Ontario
Fiona McNeill, McMaster University
Ranjana Bird, University of Windsor
Greg Weiler, Fanshawe College
Darren Lawless, Sheridan College
George Dixon, University of Waterloo
Ian Brindle, Brock University
Bill Smith, University of Ontario Institute of Technology
Michael Siu, York University
Rui Wang, Lakehead University
Patrice Sawyer, Laurentian University
Neil Emery, Trent University
Sara Diamond, Ontario College of Art & Design
John McCormick, Perimeter Institute
Murat Tuncali, Nipissing University
Mark Daley (Researcher Representative), The University of Western Ontario
Dave Frederickson, HP Canada
Hans Tuenter, Ontario Power Generation
Robert Little, Altair Engineering Canada

## Site Leaders

Thomas Wolf, Brock University
Tony Haworth/Lianne Wong, Fanshawe College
Stephen Tullis/James Wadsley, McMaster University
Ed Sykes, Sheridan College
Steve Crawford, University of Guelph
Lennaert van Veen, University of Ontario Institute of Technology
Lilia Krivodonova, University of Waterloo
Mark Daley, The University of Western Ontario
Gordon Drake, University of Windsor
Ilias Kotsireas, Wilfrid Laurier University
Michael Haslam, York University
Apichart Linhananta, Lakehead University
Kalpdrum Passi, Laurentian University
Bill Atkinson, Trent University
Michael Owen, Ontario College of Art & Design
John McCormick, Perimeter Institute

## SHARCNET Funding

Canada Foundation for Innovation (CFI)
Ontario Innovation Trust (OIT)
Ontario Research Fund (ORF)

## Private Sector Partners

Bell Canada • www.bell.ca
Hewlett Packard • www.hp.com
Nortel Networks • www.nortel.com
Optical Regional Advanced Network of Ontario (ORANO) • www.orion.on.ca
Platform Computing • www.platform.com
Quadrics Supercomputing World • www.quadrics.com
Silicon Graphics • www.sgi.com

## Academic Partner Institutions

The University of Western Ontario
University of Guelph
McMaster University
Wilfrid Laurier University
University of Windsor
Fanshawe College
Sheridan College
University of Waterloo
Brock University
University of Ontario Institute of Technology
York University
Lakehead University
Laurentian University
Trent University
Ontario College of Art & Design
Perimeter Institute
Nipissing University

SHARCNET™

The University of Western Ontario
Western Science Centre
London, ON Canada
N6A 5B7

T. (519) 661-4000
F. (519) 850-2500
E. info@sharcnet.ca
www.sharcnet.ca