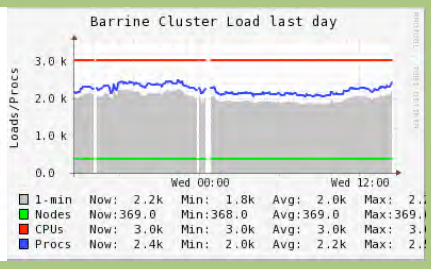
Research Computing @ UQ



5 November 2014 Issue 4

The Research Computing Centre supports collaboration to facilitate discoveries in science and engineering, humanities, and social sciences, through advanced computation, data analysis and other digital research tools.

Message from the Director

Welcome to the third newsletter for 2014. It has been an extremely busy year. A good deal of effort has been invested in building the Queensland research cloud node, together with colleagues from QCIF. Called QRIScloud, this Federal Government funded infrastructure provides a flexible compute and data cloud for researchers. We are actively working with research communities to help them move applications to the cloud and to store their data.

We have also been progressing the procurement of a new supercomputer for data intensive science, called FlashLite. At this stage we are in the final stages of selecting a vendor, and hope to announce more details in the very near future. This machine will support researchers from UQ, QUT, Griffith, UTS and Monash, as well as the CSIRO. It will be jointly managed by QCIF, and will focus on applications with a large memory footprint and high IO loads. More details will be announced in the new year.

This year we again ran a very successful seminar program, with a number of high profile speakers from the US beamed in via high definition video. These covered a range of topics from specific applications through to eResearch technologies of interest. We are currently renovating our virtual seminar room in the Axon building, and this will provide an even better space. I encourage you to keep an eye out for seminars in 2015, and I am open to suggestions from researchers on possible speakers. Please contact me if you you would like us to host a seminar for your unit.

Congratulations to all those who recently won ARC grants. We hope that if there is a need for advanced computing in your research that you won't hesitate to contact us for advice and assistance.

Since this is the last newsletter for 2014, I wish all of our readers a happy and relaxing holiday period, and I look forward to working with you in 2015.

International Conference on Computational Science

Professor Peter Sloot

RCC recently organised and ran the 2014 International Conference on Computational Science in Cairns. ICCS is an annual conference that brings together researchers and scientists from mathematics and computer science. It concerns basic computing disciplines, researchers from various application areas who are pioneering computational methods in sciences such as physics, chemistry, life sciences, and engineering, as well as in arts and humanitarian fields. The conference addresses problems and solutions in the area, to identify new issues, and to shape future directions for research.

ICCS was held in Cairns this year, the fourteenth in this series of highly successful conferences. ICCS is well known for its excellent line up of keynote speakers. The keynotes for 2014 were:

- **Professor Vassil Alexandrov**, ICREA Research Professor in Computational Science, Barcelona Supercomputing Centre, Spain
- **Professor Dr Luis Bettencourt**, Santa Fe Institute, New Mexico, USA
- **Professor Professor Peter T. Cummings**, Department of Chemical and Biomolecular Engineering, Vanderbilt University, USA
- **Dan Fay**, Director - Earth, Energy, and Environment Microsoft External Research, Microsoft
- **Dr Warren Kaplan**, Garvan Institute of Medical Research, Sydney, Australia
- **Professor Bob Pressey**, Australian Research Council Centre of Excellence for Coral Reef Studies, James Cook University, Australia
- **Professor Mark Ragan**, Institute for Molecular Bioscience, The University of Queensland, Australia

Professor Mark Ragan

The theme for ICCS 2014 was **Big Data meets Computational Science** to mark the increasing importance of data-intensive science. In order to extract meaning from the exponentially increasing amounts of data being gathered, it is imperative to both apply current computational science techniques to data sets, and to develop new processes and algorithms. This conference was a unique event focusing on recent developments in: data intensive science for diverse areas of science; scalable scientific algorithms; advanced software tools; computational grids; advanced numerical methods; and novel application areas. These innovative novel models, algorithms and tools drive new science through efficient application in areas such as physical systems, computational and systems biology, environmental systems, finance, and others.

RCC Director on SigHPC board

David Abramson is one of two new Advisory Board members appointed by SIGHPC, the Association for Computing Machinery (ACM) Special Interest Group on High Performance Computing.

SIGHPC is the first international group within a major professional society that is devoted exclusively to the needs of students, faculty, and practitioners in high performance computing. SIGHPC's mission is to help spread the use of HPC, help raise the standards of the profession, and help ensure a rich and rewarding career for people involved in the field.

The HPC community includes an incredibly wide range of individuals and organisations engaged in

studying, developing, teaching, and supporting high performance computers, storage, networking, and software. It also includes those who are working to apply HPC across a broad spectrum of scientific, engineering, and business domains. More details are available at http://www.sighpc.org/

Inaugural RCC External Advisory Committee meeting

The inaugural RCC EAC was held in June this year. This committee assists the Director in formulating overall strategic direction of RCC. A key objective is to establish RCC at the leading edge of best practice. Through its members, the EAC provides links to external organisations of both national and international significance. The committee also explores both national and international funding opportunities.

The current members of the EAC are:

- Professor Anton Middelberg, Acting DVC(R)
- Professor David Abramson, Director RCC
- Professor Paul Bonnington, Monash University
- Mr Rob Cook, CEO QCIF
- Professor David DeRoure, Director Oxford University e-Research Centre
- Professor Fran Berman, Chair US Research Data Alliance, Rensselaer Polytechnic Institute
- Professor Mike Norman, Director, San Diego Supercomputing Center,
- Professor Dieter Kranzmuller, Director at Leibniz Supercomputing Centre (LRZ) of the Bavarian Academy of Sciences and Humanities


Professor Anton Middelberg

At the June meeting the committee formally accepted the RCC strategic plan.

RCC supports first UQ Software Carpentry bootcamp

The inaugural UQ Software Carpentry bootcamp was held on 30-31 July. Bootcamps are the brainchild of Software Carpentry (http://software-carpentry.org/), a volunteer organisation founded to teach basic software skills to researchers in science, engineering, and medicine. Founded in 1998, Software Carpentry is now part of the Mozilla Science Lab (http://mozillascience.org/), which aims to transform the way science is done by teaching researchers basic skills.

The camp proved very popular with the forty or so researchers who attended. CSIRO, UQ Institutes - IMB, AIBN and QBI were represented, as well as attendees from the Mater Medical Research Institute, QIMR Berghofer, and interstate attendees from Sydney and Canberra. It was pleasing to note that 38% of the attendees were women. There was also a considerable waitlist for those wanting to attend.

Researchers were taught Python and other software and computing skills. A number of those who attended have already gone on to request cloud

resources such as virtual machines, or to use tools such as the Genomics Virtual Lab or HPC.

The bootcamp trainers were Philipp Bayer, a PhD candidate in the UQ School of Animal and Food

Sciences, Damien Irving, a PhD candidate from the University of Melbourne, and Tim McNamara from New Zealand eScience Infrastructure. They were assisted by five 'floorwalkers' expert in Python, including Tim Dettrick from the eResearch Lab in the School of ITEE, and two volunteers from RedHat in Brisbane.

The increase in researcher skills and the uptake of resources are two positive outcomes for the RCC as well as for the researchers attending.

Announcing our first RCC Fellow

The RCC is pleased to announce Associate Professor Lutz Gross from the School of Earth Sciences as its first Centre Fellow.


Lutz Gross

An important part of the RCC mission is to facilitate and promote the application of computational infrastructure (eResearch) to all fields of research and to accelerate the development of the necessary skills amongst researchers.

RCC Fellowships are awarded to outstanding UQ academic staff who have made important computational advances in their chosen fields. As RCC provides a platform for translational research, RCC forms a link between prototype research software developed by a Fellow, and trials of such systems by other groups.

RCC Fellows work together with the Centre to promote computation in university research. They engage with staff and particularly with students to explore and develop exciting new techniques and methods with impact across disciplines.

Dr Nick's Image Clinic

With some of the most advanced microscopes in the southern hemisphere, UQ researchers are generating massive amounts of imaging data. This data will help them understand processes such as organ and tissue development in health and disease, invasion of human cells by pathogens, the action of medicinal drugs, and the functions of the tens of thousands of proteins found in the genomic revolution. To fully realise the value of these advanced facilities and exciting new data sources, there is a critical need for expert advice on the most recent techniques in bio-image analysis and pipeline creation. Researchers also need to leverage the power of mathematics to conduct deeper and more sophisticated analyses of experiments.

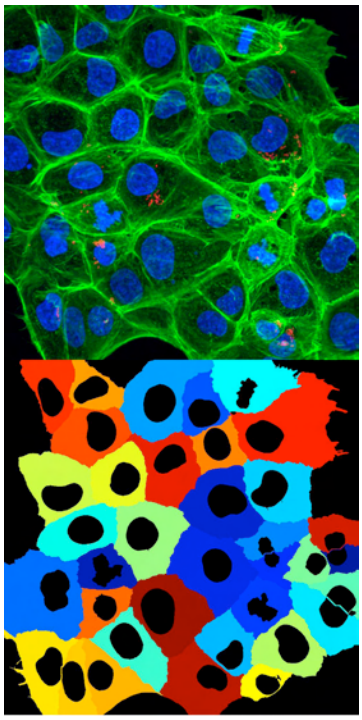

Dr Nick Hamilton

Beginning in April this year, Dr Nick Hamilton, a recent co-appointment between RCC and the Institute for Molecular Bioscience, has been running regular drop-in image clinics to provide friendly advice, training and expertise in areas such as bio-image analysis, quantification, automated image classification, mathematics, modelling, analysis pipelines, data visualisation and computational methods.

A wide range of problems have been brought to the clinic from the simple: "How can I automate counting these dot-like structures in these cell images?" to the complex: "How can we mathematically model the activation and suppression network of these proteins involved in cell adhesion and cancer?" While many problems can be solved on-the-spot with a simple analysis protocol or automation script – usually saving researchers many laborious hours of manual analysis – an initial consultation will often lead to longer term discussions. In such collaborations

experimental and analysis protocols might be developed iteratively to optimise the information extracted from imaging, or a mathematical model might be created from the data which then makes predictions that need to be tested experimentally.

The clinics have proved very popular with researchers, with some 60 consultations over 15 image clinics. While primarily accessed by researchers at IMB, the clinic has also provided advice to members of other institutes and departments at UQ as well as researchers from CSIRO and Queensland Health.



Galaxy of Tools on the Genomics Virtual Laboratory

The Genomics Virtual Laboratory's Galaxy-Qld (galaxy-qld.genome.edu.au) is a production-scale server loaded with genome indices and popular tools for use by genomics researchers, biologists and bioinformaticians. Galaxy is particularly useful for the analysis of Next Generation Sequencing data and can be used for alignments, differential gene expression, variant calling and de novo assembly.

The GVL offers self-paced tutorials which are pre-configured to run on the Galaxy Tutorial server (galaxy-tut.genome.edu.au). Workshops on these topics are run every few weeks by the GVL's partner, QFAB. See:http://www.qfab.org/ for details.

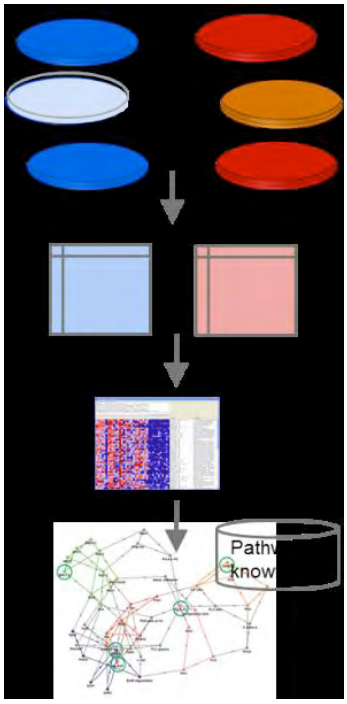
Galaxy-QLD is gaining popularity amongst UQ researchers with approximately 40 user registrations, and is running more than 1,500 jobs a month.

The cluster consists of a head node that manages the Web-based user interface and distributes jobs to the cluster, and 4 worker nodes each of 16 cores and 64GB RAM. The image below from our CloudMan management console displays a stable head node, three very busy worker nodes, and one worker node still with capacity. This cluster is running on QRIScloud, managed by QCIF http://www.qcif.edu.au/ and operated by RCC.

QRIScloud is part of the federally funded NeCTAR research cloud, which provides resources to Australian researchers. Any researcher at UQ can get a small (default size) GVL Analysis platform by applying for the resources on the NeCTAR cloud (https://dashboard.rc.nectar.org.au/). Researchers who want to go bigger can apply for enhanced allocations.

The most popular tools in use on Galaxy-QLD are Tophat, BWA and Bowtie (alignment); FastQC (quality control); Trimmomatic, Fastq_groomer and Picard tools (manipulation and processing of sequence data files); Cuffdiff (differential gene expression); Trinity (de novo transcript assembly); and Velvet (de novo genome assembly).

An interface to Blast has recently been added to Galaxy-QLD. Researchers get the usability of the Galaxy interface and the ability to integrate Blast into a workflow; with the power of Blast executing on RCC's Barrine NCI Special Facility for Bioinformatics computer.



The Queensland GVL also allows researchers access to use RStudio. For this, users need to register for an account.

To help people use Galaxy and the GVL, the RCC offers the services of biologist Dr Igor Makunin. Igor is a skilled Galaxy user and may be able to assist genomics research projects with designing a Next Generation Sequence analysis, or with specific questions about some of the popular bioinformatics tools.


Dr Igor Makunin

Big Data Sessions a highlight of 2014 Winter School

The 11th Annual Winter School in Mathematical and Computational Biology was held 7-11 July at the Queensland Biosciences Precinct with a strong program of six international, 15 national and 15 local speakers over the week.

A total of 227 participants attended with sponsorship enabling the award of 19 travel bursaries to national and international students. A particular highlight was the keynote address by Harvard Professor Johan Quackenbush on "Taming the Big Data Dragon" in a session organised by Dr Nick Hamilton. Feedback has been hugely positive with one participant commenting "it is essential for every biologist to go to Winter School!" Video of Professor Quackenbush's presentation and many others from the School are now available at http://bioinformatics.org.au/ws14/program/


Professor Quackenbush

Condolences

The centre recently lost two colleagues and collaborators, A/Professor Martin Lackmann and Dr Anouchka Mihaylova.

Vale Martin Lackmann

Associate Professor Lackmann, a biochemist at Monash University, who was instrumental in leading the development of several drugs used in clinical trials to treat cancer, passed away suddenly on 22 May, 2014. Martin was a great friend and colleague.

From the earliest years of his career, Martin was committed to ensuring that his research made the arduous journey from the laboratory to the wider world, where it could make a positive impact on people's lives.


Professor Lackmann

Granted a number of patents for his discoveries, his major contribution to the fight against cancer was the development of a drug KB004. This drug is currently in Phase II trials in leukaemia patients.

Professor Lackmann leaves a legacy of additional clinical trials and the prospect of better health for patients with brain tumours and mesothelioma.

The pioneering work of Martin Lackmann and his passion for translational research to improve the lives of others will continue through his staff and colleagues.

Vale Anouchka Mihaylova

It was with great sadness that we learned of the sudden and tragic passing of our friend and colleague Dr Anouchka Mihaylova, a project

scientist in the Department of Bioengineering at the UCSD. Anouchka and her husband were out walking near their home on May 17, 2014 when she was struck by a hit-and-run driver.


Dr Mihaylova

Anouchka worked in the area of cardiac science, and collaborated with Centre staff developing a range of computational models for cardiac ventricular cells. She visited RCC only in May 2014. A regular mentor to students, Anouchka supervised a number of UCSD Bio-engineering PRIME students, the most recent of whom visited in June this year.

Need a Barrine account?

Access to the UQ HPC resources is currently managed through the HPC unit. UQ staff and RHD students should do the following:

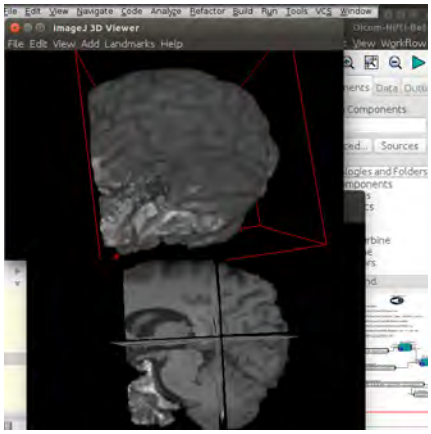
- send an email with the subject "HPC account request - with UQ login" to rcc-support@uq.edu.au
- include your UQ username, contact details and short description of your project if you are an RHD
- a ticket will be generated in the RCC Support system
- your access details will be confirmed via email.

UCSD PRIME students

RCC recently hosted three students from the University of California San Diego for their PRIME internships. PRIME students are located in a number of sites around the Pacific Rim and undertake an 8 week internship. Here are two of their reports.

Kepler Workflows for Neuroimaging Tools and 3D Image Generation, Matthew Schwegler

Advances in imaging technologies offer the promise of understanding the functions of healthy, diseased and abnormal brains. However, analysing these images requires the use of multiple tools often in an increasing number of steps (i.e., a workflow). The goal of this project is to extend the functionality of the Kepler Workflow system by integrating the workflow with tools used within the neuroimaging community. To accomplish this, we worked with researchers at the neuroimaging labs on the UQ campus to understand their work and research needs better. We then mapped the user needs into the constraints of the Kepler environment by integrating existing neuroimaging tools. We also developed a new 3D module to render images natively within the Kepler Workflow system.

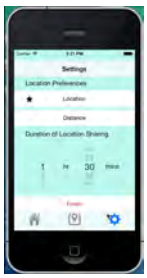


Many of the imaging tools integrated into Kepler came from FSL, a comprehensive library of neuroimaging tools, which includes the Brain Extraction Tool (BET) and Linear Image Registration Tool (FLIRT). Additionally, we developed tools for converting between the two key MRI file types, NIFTI and DICOM. With these tools functioning on MRI images within Kepler, the next step was to create a means for displaying the manipulated images in a meaningful way. This task was accomplished using ImageJ, the open-source 3D imaging API developed by the NIH. Using this API, we created a native actor (i.e., processing component) within Kepler that displays 3D images representing a new functionality inside of Kepler.

The actor was extended to work on a variety of file types such as DICOM, NIFTI, TIFF, etc. Thus the newly created, functioning Kepler workflow was able to take a series of DICOM files, extract the brain image, conduct a linear regression to the norm, and display the final product to a user.

Hoang Nguyen, a former MURPA Student who is now a UQ PhD student, developed a Web portal version of Kepler, extending it from its local running environment to a more accessible Web interface. A virtual machine of the new workflow was developed and executed in the Kepler portal. Extending the code converter to include converting NIFTI to MINC file types allowed the Web portal to display the output images of the new Kepler tools.

Smartphone App and GPS Data to Help Social Groups Manage Their Meet-ups in Real Time, Katerina Zorko



Social event planning programs are becoming increasingly popular among smartphone users and we envision improving location-sharing services specifically through our own event and location application called “Rendezvous.” The goal of this project is to develop an app that will allow users to create events and, for a period of time, it will provide access to a map displaying invited friends’ locations. This new capability can curtail texting while driving since members of an event can see each friend’s estimated arrival time and drivers will not need to send updates continually. With these arrival times available, the app can also allow users to run a few errands comfortably—rather than idly waiting for guests.

The Rendezvous app differs from other location-sharing apps in that each member of an event can specify how long they want to share their location, and each member chooses who they want to share their location with, as opposed to broadcasting their location to everybody using the app. Additionally, one of the project’s biggest goals is to protect members’ privacy. Rather than storing their phone numbers on a server, the app will assign each member an ID number to be stored instead. As much as possible, the project aims to limit and monitor the amount of personal information that is sent to the server and then shared with other members.

To date, basic design of the app has been completed, including transitioning from various screens of the app, creating new events through the user interface as well as on the server, choosing location-sharing settings, and sending notifications.

Meet RCC: Marlies Hankel

Marlies is one of our team of embedded e-Research analysis. She is a joint appointment between RCC and the Australian Institute for Bioengineering and Nanotechnology (AIBN). Originally from Germany, she studied mathematics at the University of Darmstadt and obtained the equivalent of a Master in 1997. She then moved to the University of Bristol, UK, for her PhD in theoretical chemistry. After a two year postdoc in Manchester she moved to Australia in 2004 to work at UQ, first as a postdoc and now as a research fellow. Here she managed her research group's HPC cluster while still working on her research projects and has continued to manage the group's and now the AIBN's HPC clusters. She has recently assisted in the purchasing of a new Intel Phi based cluster in the AIBN to upgrade the current facilities. Her skills are in high performance computing, computational nanotechnology and quantum & molecular dynamics calculations. She is currently working with Professor Lianzhou Wang from the AIBN and the School of Engineering on modelling of anode and cathode materials for the next generation of Lithium Ion batteries.