The Sydney Informatics Hub presents

Accelerated Computing for Innovation Conference

Friday 28 September 2018

THE UNIVERSITY OF SYDNEY
—
Core Research Facilities
—
Sydney Informatics Hub

Join us to celebrate innovation driven by cutting-edge tech, big data, machine learning and AI. Listen to fascinating talks from science and industry about data science and technology powering precision medicine, wildlife and crop conservation, virtual earth explorations, rock and shark spotting, education and language tech, crime prediction, and more! Leave full of ideas, connected, and motivated.

Free registration

informatics.sydney.edu.au/hpc_conference/

Location

ABS Lecture Theatre 1130
Abercrombie Building (H70)
cnr Abercrombie St and Codrington St
University of Sydney

Program

8.45am - 9.00am Registration
9.00am - 9.10am Welcome
9.10am - 11.15am Life, Evolution and Health
11.15am - 11.30am Morning tea
11.30am - 12.30pm National Collaborative Infrastructure
12.30pm - 1.15pm Lunch
1.15pm - 3.15pm Earth, Environment and Space
3.15pm - 3.30pm Afternoon tea
3.30pm - 5.35pm Humans and Society
5.35pm - 5.45pm Closing Remarks

More information

Sydney Informatics Hub
T +61 2 8627 6286
E sih.info@sydney.edu.au

Accelerated Computing for Innovation Conference

Friday 28 September 2018

ABS Lecture Theatre 1130
Abercrombie Building (H70)
cnr Abercrombie St & Codrington St
University of Sydney
informatics.sydney.edu.au/hpc_conference/

THE UNIVERSITY OF SYDNEY
Core Research Facilities
Sydney Informatics Hub

Program

Time	Session
8.45am	**Registration**
9.00am	**Opening Address** **Jeremy Hammond**, USyd
9.10am	**Life, Evolution and Health** **Aidan O'Brien**, CSIRO - Life science research at scale **Mark Pinese**, Garvan - Computing the genomics of good health **Carolyn Hogg**, USyd - The future is here - the benefits of genomic tools in translocations **John Sebastian Eden**, WIMR, USyd - Revealing the hidden world of viruses, one Giga-base pair at a time **Emily Remnant**, USyd - Using big data to find small things inside bees
11.15am	**Morning tea (provided)**
11.30am	**National Collaborative Infrastructure** **Andrew Lonie**, UniMelb - Planning digital research infrastructure for biosciences **Christopher Harris**, Pawsey - Enabling Petascale Science with the Pawsey Supercomputing Centre **Ben Evans**, NCI - Current challenges in cross-disciplinary computational and data integration
12.30pm	**Lunch (provided)**
1.15pm	**Earth, Environment and Space** **Geraint Lewis**, USyd - What's happening in the dark-side of the universe? **Dietmar Muller**, USyd - Understanding Earth System Evolution - connecting Deep to Surface Processes **Caroline Foster**, USyd - Galaxy formation and evolution **Cormac Purcell**, MQ - Sharks, tombs and cosmic bubbles: how machine-learning spans diverse scientific fields
3.15pm	**Afternoon tea (provided)**
3.30pm	**Humans and Society** **Kai Riemer and Sandra Peter**, USyd - AI - History, hype and hysteria **Simon Hammond**, Appen - Understanding speech and text in low-resource languages **Bing Ong**, Daisee - Democratising AI for businesses **Sonya Corcoran**, Digital Innovation - Who is making what and how? **Roman Marchant**, USyd - Understanding and preventing crime using Data Science
5.35pm	**Closing Remarks** **Simon Ringer**, USyd

THE UNIVERSITY OF SYDNEY
Core Research Facilities
Sydney Informatics Hub

Accelerated Computing for Innovation Conference

Friday 28 September 2018
informatics.sydney.edu.au/hpc_conference/

ABS Lecture Theatre 1130
Abercrombie Building (H70)
cnr Abercrombie St & Codrington St
University of Sydney

Conference Abstracts

Opening Address
9.00am – 9.10am

Opening address
9.00am – 9.10am

Jeremy Hammond, University of Sydney

Dr Jeremy Hammond is the Director, Strategic Ventures at the University of Sydney, where he leads university-wide transformation initiatives. He is a linguist by trade, with a PhD from the Max Planck Institute for Psycholinguistics specialising in field experiments on indigenous languages in Southern Vanuatu and has experience in developing capabilities in high performance computing, informatics, and data services. These days, both his coding and Austronesian languages are rusty with his focus now on transforming the University for the next generation of researchers and students.

Life, Evolution and Health
9.10am – 11.15am

Life science research at scale
9.10am – 9.35am

Aiden O'Brien, CSIRO

Genomic data is outpacing traditional big data disciplines, producing more information than astronomy, Twitter, and YouTube combined. As such, genomic research has leapfrogged to the forefront of big data, requiring novel cloud solutions with artificial intelligence and machine learning to generate useful insights. This talk therefore showcases how we find disease genes responsible for ALS using VariantSpark, a random forest implementation built on Spark to deal with the 80 million dimensions in genomic data. Next, this talk outlines how we use a serverless architecture to translate these insights into the clinical practice by providing a decision support framework for clinicians. Finally, the talk touches on how to evolve serverless architecture more efficiently through a hypothesis-driven approach to DevOps and how we keep data and functions secure in a serverless environment.

Aidan O'Brien is a PhD candidate holding a joint position with Australia's national science agency, the CSIRO, and the Australian National University. He is tasked to develop sophisticated machine learning models for accurate genome editing as needed in gene therapy for human health. He graduated from the University of Queensland with a Bachelor of Biotechnology (1st class honours) in 2013 where he developed GT-Scan, a CRISPR target predictor. Aidan then started at the CSIRO with the transformational bioinformatics team, where he developed VariantSpark, which applies machine learning to big genomic data. He has an enthusiasm for serverless (FaaS) and hypothesis-driven architecture design, which he spoke about at YOW! Data 2018. Aidan has 5 journal publications (3 first author) with 86 citations (h-index 3) and has attracted $180K in funding to date as AI. He received the "Best student and postdoc" award at CSIRO in 2015.

Computing the genomics of good health
9.35am – 10.00am

Mark Pinese, Garvan Institute of Medical Research

Dr Mark Pinese is a Senior Research Officer in the Garvan's Genomic Cancer Medicine Lab where he is the lead analyst for the Medical Genome Reference Bank and conducts research into the genetic basis of cancer risk.

The future is here - the benefits of genomic tools in translocations
10.00am – 10.25am

Carolyn Hogg, University of Sydney

Since the publication of the human genome in 2001, the development of genetic and genomic tools for conservation management have come forward in leaps and bounds. While genetic data used to be considered an academic "nice to have", it is fast becoming an essential item in the translocation of

endangered species. IUCN guidelines recommend that translocations should occur when the threatening process has been reduced or removed. However, when infectious diseases are a threatening process, alternative management strategies are required. Devil facial tumour disease has decimated populations by 80% over the past 20 years. Modelling indicates that this genetically depauperate species will succumb to small population pressures, such as inbreeding depression, if augmentation of wild populations do not occur. We will show how genomic data has not only informed management practices of the Tasmanian devil metapopulation but how it is being used for translocations in the wild. Since 2011, we have implemented a strict genetic and demographic management strategy for devils, both in the wild and in captivity. Our strategy is directly informed by our genomic work, and we will describe the benefits, our successes and our failures in using cost-effective genetic tools in the augmentation of wild populations of an iconic Australian species.

Dr Carolyn Hogg is the Research Manager for the Australasian Wildlife Genomics Group in the Faculty of Science, University of Sydney. Carolyn is a conservation biologist who has been working with threatened species for over twenty years both in Australia and overseas. Carolyn has been working closely with the Save the Tasmanian Devil Program for over eight years and with other threatened species more recently. Through her partnerships with both her academic and conservation colleagues Carolyn has developed better tools and technologies to integrate molecular genetics into real-time conservation management decisions.

Revealing the hidden world of viruses, one Giga-base pair at a time.
10.25am – 10.50am

John Sebastian Eden, Westmead Institute for Medical Research

The field of pathogen genomics has been transformed by new sequencing technologies. From practically any sample, billions of sequence reads can be generated that provide an unbiased snapshot of what viruses, bacteria, fungi, and other parasites, might be present. Meta-transcriptomics or bulk RNA sequencing has proven to be particularly powerful for viral discovery and our research highlights the richness of this data. However, the ease at which data can be produced is contrasted sharply against the computational and analytical challenges in handling the volumes of data as well as being able to accurately describe genomes which are incredibly diverse. Here, I will present data from our recent metagenomic studies and outline how we utilise high-performing computing to address some of the previously mentioned challenges.

Dr John-Sebastian Eden is a Research Scientist in Bioinformatics and Genomics at the Westmead Institute for Medical Research and Senior Research Fellow in the Sydney Medical School. Previously, he was an NHMRC Early Career Fellow working with Prof Eddie Holmes at the University of Sydney. His research uses pathogen genomics and phylodynamics to better understand the sources and mechanisms of how pathogens emerge, spread and evolve in populations. Recent work includes the development of novel assays for sequencing individual HIV proviruses to explore the latent reservoir and the application of "meta-transcriptomics" to pathogen discovery in both human and wildlife disease investigations.

Using big data to find small things inside bees: Discovering new honey bee viruses with metatranscriptomics
10.50am – 11.15am

Emily Remnant

From hunter-gatherers to air-conditioned supermarket aisles, technology has revolutionised the way we access food. Industrialised agricultural practices rely in no small part on pollination services from honey bees, arguably the unsung heroes of every meal. Pollination from honey bees improves the quality and diversity of the food we eat. However, increasing demand for pollination has seen dramatic changes in the management of honey bees. Worldwide, we have experienced a decline in the health of bees due to a cocktail of novel stressors such as pesticides and diseases- in particular, viruses spread by parasitic mites- putting global food supplies at risk.

Genomic data has immense potential to help us improve the health of honey bees, from breeding healthier bees, to better understanding the spread and management of disease. I will outline how metatranscriptomics has contributed significantly to our understanding of virus diversity in honey bees, by enabling the discovery of completely new viruses, to identifying strains of highly pathogenic viruses, and how genomics offers hope in the form of new solutions to preventing viral disease in our most important agricultural pollinators.

Dr Emily Remnant is a geneticist and evolutionary biologist in the Social Insects Lab, University of Sydney investigating how insects respond to changes in their environment. She use genetics and genomics to study genes involved in resistance to chemical insecticides, and to investigate the impact of viral diseases in beneficial insects like honey bees on crop pollination.

National Collaborative Infrastructure
11.30am – 1.15pm

Planning digital research infrastructure for biosciences
11.30am – 11.50am

Andrew Lonie, University of Melbourne

Some research disciplines have advanced, multi-year community-sourced roadmaps for digital infrastructure in Australia; unfortunately, biosciences isn't one of those disciplines.
Bioplatforms Australia, through the EMBL Australia Bioinformatics Resource and associates, is working to define a national 'bioinformatics commons' – national research infrastructure to support widespread access to the critical data, tools, compute, techniques and training required for high quality analysis of digital data in biosciences. This means understanding the different types of researchers using and doing bioinformatics, what their needs are, what is happening globally, and how we can work together as a country to benefit locally and engage internationally. If done properly, we believe this work will inform and influence national infrastructure investments in this space long term, particularly the Australian Research Data Commons (https://www.ands-nectar-rds.org.au/).
In this talk I will provide an overview of the bioinformatics commons project and discuss the process and progress to date, including engaging with the bioinformatics community and future plans.

Andrew Lonie is Director of Melbourne Bioinformatics and an associate professor at the Faculty of Medicine, Dentistry and Health Sciences at the University of Melbourne, where he coordinates the MSc (Bioinformatics). Andrew directs a group of bioinformaticians, computational biologists and HPC specialists to collaborate with and support life sciences researchers in a variety of research projects within the Parkville research precinct and across Australia.

Enabling Petascale Science with the Pawsey Supercomputing Centre
11.50am – 12.10pm

Christopher Harris, Pawsey Supercomputing Centre

The Pawsey Supercomputing Centre encourages and energises research using supercomputing, large scale data storage and visualisation. We provide facilities and expertise to research, education and industrial communities across Australia. Application areas include nanotechnology, radio astronomy, high energy physics, medical research, mining and petroleum, architecture and construction, multimedia, and urban planning, and many more.
This presentation will provide an overview of Pawsey facilities that are available nationally to Australian researchers, including details of how to get access, as well as showcase some of the currently supported research projects.

Dr Christopher Harris is a Senior Supercomputing Specialist at the Pawsey Supercomputing Centre. He provides assistance and expertise to researchers, enabling computational workflows to make effective use of Australia's national petascale HPC facilities.

Current challenges in cross-disciplinary computational and data integration
12.10pm – 12.30pm

Ben Evans, National Computational Infrastructure

In the next few years, there will be a major increase in computational power - through upgraded HPC systems or further uptake of cloud-based platforms. At the same time, an enormous amount of new digital data will come on-line from many domains. However, the two do not simply come together, and there will be barriers for data-intensive science and the full realisation of the potential of deep learning, AI, and other new computationally-intensive technologies. In many cases, the data will need to be better organised to make it more tractable to process at-scale, and to make it programmatically accessible for a range of use-cases.

Over the last several years, NCI has been focused on improving computational access to some major national reference geospatial datasets. Beyond the NCI environment, there has also been significant update via remote access via data services, including server-side data processing - utilising the colocation of data and computational processing power. With the data too big or too complex to move, we are now in the post-download era.

The next challenge is to enable these techniques to be usable across multiple domains: this will necessitate an increasing focus on "FAIR data" principles - Findable, Accessible, Interoperable and Reusable. FAIR is underpinned by the concerted efforts that are happening internationally to develop community agreed standards that enable seamless programmatic access to data in high-performance environments across multiple domains. While this places additional requirements on the suppliers of both the data and metadata, the result is that data can be even more accessible - for primary use, secondary use, and citability through publication processes.

Dr Ben Evans is the NCI Associate Director, Research Engagements and Initiatives. He is an expert in HPC, data management and software development. For the last several years he has particularly focused on national priority areas of research at NCI: weather, climate, earth systems, environmental sciences and geophysics.

Earth, Environment and Space
1.15pm – 3.15pm

What's happening in the dark-side of the universe?
1.15pm – 1.45pm

Geraint Lewis, University of Sydney

Professor Geraint F Lewis was born and raised in Old South Wales, and completed degrees in London and Cambridge before undertaking research in Canada and the US. He moved to Australia in 2000 and is now professor of astrophysics at the Sydney Institute for Astronomy at the University of Sydney. His research focuses on cosmology, gravitational lensing and galactic cannibalism, all with the goal of unravelling the dark-side of the universe. He is co-author of A Fortunate Universe: Life in a finely tuned cosmos, which examines why the physical properties of the universe appear to be just right for complexity and life.

Understanding Earth System Evolution-connecting Deep to Surface Processes
1.45pm – 2.15pm

Dietmar Müller, University of Sydney

Our planet is of a composition and a location in space that has resulted in a thermal and structural evolution that is unique in the solar system, forming a resource-rich, oxygenated habitable Earth. Our civilization is predicated on stable climate and coastlines, yet the geological record reveals numerous episodes of enormous change, innovation, radiation and extinction. We need to understand the underlying mechanisms and feedbacks to predict possible future paths of the Earth. How do the evolution of continents and climates affect ocean circulation, shifts in coastlines, rainfall, rock weathering and soil formation or erosion? How long may it take for geological forces, as opposed to humans, to dominate Earth's surface evolution again in the future? Human civilisation equally depends on a range of metal resources to build renewable energy infrastructure. To identify portions of the Earth's crust rich in these resources, we need to understand how bursts in mineralisation are related to the interplay between the evolution of the deep Earth, the crust and surface environments. Recent advances in computing power, software development and global database development enable rapid advances in modelling the interconnectivity of Earth system processes, transforming our ability to understand past and possible future evolutionary paths of our planet.

Dietmar Müller received his undergraduate degree from the Univ. of Kiel, Germany, and his PhD in Earth Science from the Scripps Institution of Oceanography, La Jolla/California in 1993. After joining the University of Sydney he built the EarthByte Research Group, pursuing the collaborative development of open-source software, open-access data sets and virtual globes. The fundamental aim of the EarthByte Group is geodata synthesis through space and time, assimilating the wealth of disparate geological and geophysical data into a four-dimensional Earth model. His innovations have led to a Laureate Fellowship and a succession of industry partnerships (three ARC Linkages, two directly funded industry projects, one ARC Industry Transformation Research Hub). Applications include mineral prospectivity maps, modelling the geodynamic history of resource-rich sedimentary basins

around Australia and Southeast Asia, and understanding the drivers of long-term climate change in a paleogeographic framework. Dietmar is a fellow of the Australian Academy of Science and the American Geophysical Union.

Solving the century old challenge of measuring the true shape of galaxies
2.15pm – 2.45pm

Caroline Foster, University of Sydney

The nature of astronomical distances is such that even our nearest neighboring galaxy is too far to explore by probe. Much like shadow puppetry, astronomers see galaxies in projection. Since distinct shapes project identically in 2D, measuring the true 3D (or intrinsic) shape of galaxies is a challenging problem that is nearly a century old. I will describe the problem and show how recent technological advances are finally allowing astronomers to reliably measure the intrinsic shape of galaxies.

Dr Caroline Foster is the Astro3D Centre for Research Excellence fellow at the University of Sydney. Her research focuses on galaxy formation and evolution through the analysis of the 3D structure, dynamics and chemistry of galaxies.

Sharks, tombs and cosmic bubbles: how machine-learning spans diverse scientific fields
2.45pm – 3.15pm

Cormac Purcell, Macquarie University

Science in the 21st century is a vast undertaking and most of us at the coal-face specialise in very narrow fields. Universities love the idea of innovative collaborations across disciplines, but it is often difficult to form such teams in practice. Here I talk about my personal experience widening my horizons as an astronomer, to embrace ecology and archaeology.
Cormac Purcell a researcher and lecturer in the Department of Physics and Astronomy at Macquarie University. He uses radio telescopes to investigate the interstellar medium, with a focus on magnetism and star-formation in the Milky Way.

Humans and Society
3.30pm – 5.30pm

AI - History, hype and hysteria
3.30pm – 3.55pm

Kai Riemer and Sandra Peter, University of Sydney

Kai Riemer is the Professor of Information Technology and Organisation, Business Information Systems, Sydney Business School. He is the founder and leader of the Digital Disruption Research Group (DDRG) and convener of DISRUPT.SYDNEY™. His expertise spans the fields of digital commerce, inter-firm networking, the future of work, technology adoption, and the philosophy of technology.

Sandra Peter is the Director for Sydney Business Insights with expertise in disruptive innovations for education and the future of business.

Understanding speech and text in low-resource languages
3.55pm – 4.20pm

Simon Hammond, Appen

Simon Hammond has worked in speech technology for more than a decade, producing training corpora in over 30 different languages. He specialises in translating between teams of linguists and engineers.

The Low Resource Languages for Emergent Incidents (LORELEI) program aims to automatically detect emergencies and needs for international aid in languages from Haitian Creole to Tamil. But language technology has come to rely on huge amounts of training data for its accuracy; how can we adapt these engines to languages with few existing resources?

Democratising AI for businesses
4.20pm – 4.45pm

Bing Ong, Daisee

AI start-up, Daisee, saw a gap in the Australian market, with few specialist AI companies developing software for businesses of any size to use. It's focus is on developing software while helping companies quickly integrate customised AI capabilities. Daisee's data scientist, Bing Ong, will talk about the road to developing its first product, and how opening up AI for businesses is going to change the world in which we live, starting from the call center.

Bing is currently a data scientist at Daisee, a local start-up that specialises in artificial intelligence. Bing has over seventeen years of experience in IT, working for startups to MNCs (eg. Oracle (nearly ten years)) with experience spanning across anything to do with data – presales, system integration, project management and program development. She found her passion in data science after recently completing her studies in the Masters of Data Science and Innovation with UTS. In her new role at Daisee, she is continuing her knowledge-seeking journey by working with world-class consultants and talented data scientist to provide value to clients.

Digital Innovation - Who is making what and how?
4.45pm – 5.10pm

Sonya Corcoran, University of Sydney

Sonya Corcoran is the manager for Digital Innovation. She leads the TechLab where students and researchers can experiment with new technology and algorthims. She is also a co-founder of Changineers where she helps to design experiential learning technology.

Understanding and preventing crime using Data Science
5.10pm – 5.35pm

Roman Marchant, University of Sydney

The use of quantitative models for uncovering patterns in crime has become increasingly popular world wide. In this talk I explore the models that we have developed to further understand the ocurrance of crime, mainly focussing on the use a fully Bayesian methodology to estimate relationships between demographics, space, time and a variety of crime categories. Coupled with complex models, comes the use of high end computational infrastructure which allows to solve and infer patterns in these large quantities of data. I not only explore previous research, but will also present preliminary results in our current research for crime prevention by optimising patrolling trajectories and further understanding long term consequences of socieatal issues associated with specific crime types, such as domestic violence.

Dr Roman Marchant is a Research Fellow at the Centre for Translational Data Science. He applies data science to predict crime and to understand criminal behaviour. His work informs governmental policies and policing programs which are designed to curb domestic violence levels within the community.

Closing remarks
5.35pm – 5.45pm

Closing Remarks
5.35pm – 5.45pm

Simon Ringer

Professor Simon Ringer is the University of Sydney's Academic Director of Core Research Facilities, and leads the University's strategic planning and operations of its high-end research infrastructure initiatives. He is a Professor of Materials Science and Engineering in the School of Aerospace, Mechanical & Mechatronic Engineering, and an academic member of the Australian Centre for Microscopy & Microanalysis.

The **Accelerated Computing for Innovation Conference** is sponsored and supported by:

Our speakers are supported by: