AOTEAROA

eRangahau Research NZ 2025

Abstracts Handbook

11-13 February 2025

16th annual conference

ŌTAUTAHI CHRISTCHURCH

Many thanks to our conference sponsors:

PLATINUM and COFFEE CART CO-SPONSORS

XENON High Performance Computing | NVIDIA

GOLD SPONSORS

sempre Always Available

DELL Technologies TITANIUM PARTNER

CDC

DIGITAL science

WELCOME RECEPTION CO-SPONSORS

Hewlett Packard Enterprise | NVIDIA

BRONZE SPONSOR

ARDC

Australian Research Data Commons

OTHER SUPPORTERS

GIGABYTE™ WEKA Versity

AOTEAROA eRangahau Research NZ

A future for publicly funded eResearch infrastructure

Amber McEwen
REANNZ
amber.mcewen@reannz.co.nz

ABSTRACT

Following Minister Collin's request that REANNZ expand its remit to offer a broader range of services to support and grow Aotearoa's eResearch sector Amber McEwen, the REANNZ CEO, will provide an update on how REANNZ is planning to do more to support the sector over the years ahead.

ABOUT THE AUTHOR

Amber McEwen is a leader with over 20 years' experience in B2B service industries. She has a proven track record in generating revenues and creating brand differentiation through the development, implementation and management of services over networks ranging from telecoms to energy.
Driving collaboration at many levels (globally, cross-sector and with customers) is a strength of Amber's and is key to the success she has delivered. Throughout her career this has been evidenced at a global level while at Vodafone, and at a cross-sector level while at ESR. She says it is building these relationships and delivering outcomes that make a material difference that she finds most personally and professionally rewarding.

Aotearoa Genomic Data Repository: the journey to a production repository… and beyond!

Mik Black, University of Otago, mik.black@otago.ac.nz
Claire Rye, New Zealand eScience Infrastructure, claire.rye@nesi.org.nz
Libby Liggins, University of Auckland
Rudiger Brauning, AgResearch
Tracey Godfery, Genomics Aotearoa / University of Otago
Tanis Godwin, Genomics Aotearoa / University of Otago
Jun Huh, New Zealand eScience Infrastructure
Eirian Perkins, New Zealand eScience Infrastructure
Nathalie Giraudon, New Zealand eScience Infrastructure

ABSTRACT

The Aotearoa Genomic Data Repository (AGDR) is a joint initiative between Genomics Aotearoa and NeSI to provide a secure within-nation repository for the storage, management and sharing of non-human genomic data generated from biological and environmental samples originating in Aotearoa New Zealand. Developed following the principles of Māori Data Sovereignty, and enabling data management through kaitiakitanga, the AGDR allows iwi, hapū and whānau to effectively exercise their responsibilities as guardians over biological entities they regard as taonga.
Four years since the initial launch, the team will review the journey thus far, the design decisions and progress, lessons learnt, the response from the community both in terms of researchers submitting and accessing data, and kaitiaki carrying out their responsibilities for the taonga data. We will also present some thinking on the future directions of the repository and where we are headed over the next five years.

ABOUT THE AUTHORS

Professor Mik Black is a statistician, bioinformatician and data scientist whose research focuses on the development of methods for the analysis of genomic data, with a strong emphasis on cancer and other human diseases. A common theme is the use of techniques that allow high-dimensional and often very disparate data sets to be combined in ways that provide new insights into disease development and progression. In addition to his own work, Mik has been heavily involved in establishing national research infrastructure in high performance computing through the NZ eScience Infrastructure, and in genomics and bioinformatics through Genomics Aotearoa, where he is Co-Director and chair of the Bioinformatics Leadership Team.

Dr Claire Rye is a Product Manager at New Zealand eScience Infrastructure (NeSI) based out of the University of Auckland. She is responsible for the National Data Transfer Service and works across the Aotearoa Genomics Data Repository and Rakeiora Pathfinder projects and looking at research data management and data lifecycle more generally across NeSI. Claire holds a PhD in organic chemistry and has spent the last 11 years working in the UK in a variety of research settings.

Assessing the support for scientific claims using generative and general AI

Mark Gahegan[1], Neset Tan[1], Niket Tandon[2], David Wadden[2], Oyvind Tafjord[2], Michael Witbrock[1]
[1]University of Auckland, New Zealand, [2]Allen Institute for Artificial Intelligence, Seattle, USA
m.gahegan@auckland.ac.nz

ABSTRACT

When computers first emerged in research in the 1960s-70s, they were used to scale up analysis and modelling. Their impact was huge in some fields, but marginal in others, being limited to just one part of the research lifecycle.
This time is different. As we grasp what AI can do, or will do soon, it is clear the impact on research will be massive: much more of the research lifecycle will be affected—many more research activities we be augmented by AI.
This talk describes a new tool that can evaluate the truth of any general scientific claim against the research literature, *without any pre-training*, using existing generative AI and more established computational linguistics methods. The tool has a simple interface, can evaluate multiple perspectives and combines extracted information and reputational metrics in reaching its conclusions. We explain how it works and give examples of its use. We also discuss how tools such as this will impact the conduct of research into the future.

ABOUT THE AUTHOR

Mark Gahegan is the founding director of the Centre for e-Research and Professor in the School of Computer Science at the University of Auckland, New Zealand. Mark has wide interests at the intersection of Computer Science, Machine Learning, eScience and GIScience. This is because he can never sit still for too long and is easily distracted. He is kind of scruffy-looking. He also directs the ‘Beyond Prediction...’ national Data Science program, focussed on applying emerging Machine Learning, eScience and GenAI methods to problems in Ecology, genomics and infectious diseases.

Building a Secure Research Environment

Martin Feller[1] and Yvette Wharton[1]
[1] Waipapa Taumata Rau, University of Auckland
m.feller@aucklad.ac.nz, y.wharton@auckland.ac.nz

ABSTRACT

Supporting collaboration and innovation while protecting and managing sensitive research data in an increasingly important cybersecurity and risk landscape is challenging for universities and researchers.

Providing tools and environments that keep data safe and allow secure access for approved researchers to work collaboratively and analyse sensitive data, augmented with layers of governance and auditing, is one response to these challenges.

This presentation will discuss our activities to improve sensitive research data management practices and our work to date on building our Secure Research Environment (SRE). We will share our experiences building a SRE and outline and explore the enablers and the hurdles we have encountered.

ABOUT THE AUTHORS

Martin Feller Since moving to New Zealand Martin Feller has worked at the Centre for eResearch at the University of Auckland in various roles. He is currently leading the Platform and Services Team, which is looking after the operations of the Nectar OpenStack platform, other backend services, automation and reporting.

Yvette Wharton works at the Centre for eResearch, Waipapa Taumata Rau, University of Auckland. She is the Technical Business Lead for the Research Data Management Programme, working on the Secure Research Environment and machine-actionable Data Management Planning initiatives. http://orcid.org/0000-0002-6689-8840

Building a Skilled Research Workforce: Strengthening the National Training Landscape

François Bissey, Michael Coe, Murray Cadzow, Nisha Ghatak, Rina Hannaford, Tom Saunders, Tyler McInnes and Wes Harrell
University of Canterbury, University of Otago, New Zealand eScience Infrastructure, AgResearch, University of Auckland, Genomics Aotearoa and Manaaki Whenua Landcare Research
francois.bissey@canterbury.ac.nz, michael.coe@canterbury.ac.nz, murray.cadzow@otago.ac.nz, nisha.ghatak@nesi.org.nz, rina.hannaford@agresearch.co.nz, tom.saunders@auckland.ac.nz, tyler.mcinnes@otago.ac.nz, HarrellW@landcareresearch.co.nz

Situation:

Given the current uncertainties around budgets and increasingly limited personnel, developing a strong and skilled digital research workforce in New Zealand poses significant challenges. The need for consistent, high-quality digital training across the country has grown, particularly within Crown Research Institutes (CRIs) and other research institutions where upskilling in advanced computing and data science is essential. However, the scarcity of resources has brought forth the need to maximise impact through efficient coordination, resource sharing, and innovative approaches to training.

Task:

This session will bring together stakeholders from across the national research sector to discuss practical solutions for advancing New Zealand's training landscape. We aim to explore collaborative models that reduce duplicated efforts across institutions, increase regional support for training, and leverage NeSI's role as a national connector. Participants will share insights on how to strengthen coordination, utilise local expertise, and enhance digital competencies in research through more inclusive and accessible training initiatives.

Action:

During this Birds of a Feather session, we will facilitate a collaborative discussion on the most effective ways to support and expand New Zealand's digital skills training. Topics include:

- Approaches to upskilling university and CRI staff, while aligning training with organisational goals.

- Practical strategies for coordinating and sharing training resources across institutions.
- The unique value NeSI could bring in as a national training conduit and how to maximise their impact.
- Models for sustainable regional support in training delivery to extend reach.
- Innovative ways to do more with limited budgets and personnel, using examples from other nations and fields.

Result:

By the end of this session, we aim to develop actionable strategies and identify potential collaborative training efforts that can enhance New Zealand's research capabilities. Key takeaways will be synthesised into a report, outlining opportunities for coordinated training and resource sharing, as well as recommendations for leveraging national and regional resources more effectively.

ABOUT THE AUTHORS

Mr Michael Coe - Michael is an Assistant eResearch Consultant at University of Canterbury. Michael, along with François Bissey, assist UC researchers with software engineering, data management, HPC, digital skills training and bespoke research requests.

Dr Murray Cadzow - Murray is a Scientific Programmer within Research Teaching IT Support at the University of Otago, with a background in bioinformatics. Murray has been heavily involved in computational literacy and bioinformatic training at the University of Otago - co-organising Research Bazaar and the Otago Bioinformatics Spring School. He is certified as both a Carpentries instructor, and instructor trainer.

Dr Nisha Ghatak – Nisha is a Research Communities Advisor – Training Lead at New Zealand eScience Infrastructure. She facilitates skill development efforts nationally and supports training communities through her role as a community coordinator for The Carpentries in Aotearoa New Zealand. Nisha is also a certified Carpentries instructor and part of the Board of Directors for The Carpentries.

Dr Rina Hannaford - Rina is a data scientist at AgResearch who is involved in the organisation's emerging digital skills development programme. Rina is a certified Carpentries instructor and has a background in statistics.

Dr Tom Saunders - Tom is an Engagement Specialist in the Centre for eResearch, University of Auckland, where he coordinates the Centre's digital research skills training programme. He is a co-organiser of ResBaz Aotearoa and a certified Carpentries instructor.

Dr Tyler McInnes - Tyler is the Bioinformatics Training Coordinator for Genomics Aotearoa, where he oversees the development and implementation of a portfolio of training workshops. Tyler is a certified Carpentries instructor, and has a background in genomics, bioinformatics, and training.

Mr Wes Harrell - Wes is eResearch Support Manager at Manaaki Whenua. Wes and the rest of the Manaaki Whenua eResearch team assist with software engineering, data engineering, HPC, devops, training and other things digital.

Building the Future of the RSE-AUNZ Community: Opportunities, Strategy, and Engagement

Mercedes Randell, NeSI mercedes.mudgway@nesi.org.nz
Rowland Mosbergen, WEHI mosbergen.r@wehi.edu.au
Manodeep Sinha, ACCESS-NRI, manodeep.sinha@anu.edu.au
Peter Marendy, QCIF peter.marendy@qcif.edu.au

ABSTRACT
The Research Software Engineer Australia New Zealand (RSE-AUNZ) community has grown to 479 members, supporting professionals who are essential to the digital future of research. This presentation will provide an overview of RSE-AUNZ's current state, recent achievements, and its strategic path toward a sustainable, inclusive community. With goals to connect members through local meetups and tech talks, raise the profile of RSEs, foster diversity, and build a strong community, RSE-AUNZ aims to empower individuals and strengthen the digital research ecosystem.

We will discuss efforts to grow active engagement, including appointing local and domain-specific champions, developing resources to enhance career pathways, and seeking collaborations with other communities to create a broader network of practice. Smaller, cross-domain talks and regional meetups are in development to make participation accessible and rewarding, while plans for grant support are underway to secure lasting growth.

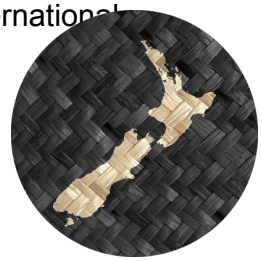
In addressing both the challenges and the future of research software engineering in Australasia, this session will demonstrate how RSE-AUNZ is helping to shape a robust, supportive community that values skill development and sustained engagement—paving the way for new digital horizons in research.

ABOUT THE AUTHORS

Mercedes Randell
Mercedes Randell is a Research Communities Advisor at New Zealand eScience Infrastructure (NeSI), where she fosters collaboration within the research software engineering community. With a background in neuroscience, her research focused on how cognitive processes such as attention and memory influence decision-making and learning. Having worked in the research community, Mercedes is a strong advocate for accessible and open science. She is passionate about supporting the development and adoption of open tools in research software. Additionally, Mercedes serves as a Steering Committee Co-Chair for RSE-AUNZ.

Rowland Mosbergen
Rowland Mosbergen is a Research Software Engineer at the Walter and Eliza Hall Institute of Medical Research (WEHI) in Melbourne, Australia. With over 25 years of experience in IT, he specialises in designing, developing, and implementing research data analytics solutions. Rowland is also a Diversity, Equity, and Inclusion advocate and writer, focusing on creating inclusive environments within the research community. He has contributed to various initiatives aimed at fostering diversity and inclusion in research software engineering. Rowland is on the steering committee for RSE-AUNZ and is the RSE-AUNZ representative on the International Council of RSE Associations.

Mandeep Sinha



Dr. Manodeep Sinha works on Australian climate change models within ACCESS-NRI at the Australian National University (ANU). Previously, he was a Senior Research Software Scientist at Swinburne University of Technology, where he worked on astrophysical simulations, including galaxy evolution and reionisation. Manodeep is also the founder of RSE-AUNZ chapter, where he promotes accessibility and collaboration in research software development across various scientific domains, including climate science and astronomy.

Peter Marendy

Dr Peter Marendy is the Head of Data and Software Solutions at the Queensland Cyber Infrastructure Foundation (QCIF), where he leads initiatives to provide advanced data services to the research community. With over 20 years of experience in software engineering across various research and innovation domains, Peter has contributed to numerous projects aimed at advancing computational informatics. Prior to his role at QCIF, Peter worked at the Commonwealth Scientific and Industrial Research Organisation (CSIRO), focusing on computational informatics. Rowland is on the steering committee for RSE-AUNZ.

Celebrating Diversity in High Performance Computing: A Spotlight on WHPC+ Australasia

Jana Makar, New Zealand eScience Infrastructure (NeSI)
Aditi Subramanya, Pawsey Supercomputing Research Centre
Emily Barker, University of Western Australia
Linda McIver, Australian Data Science Education Institute
Carina Kemp, Amazon Web Services (AWS)
Kiowa Scott-Hurley, Independent
Kerri Wait, Independent
ahoy@whpcanz.org

ABSTRACT
Visibility matters. As a Chapter of Women in High Performance Computing (WHPC), we recognise the role we play in raising the visibility and celebrating the successes of our community members. Our Australasian Chapter adds a "+" to our name to emphasise that we welcome people from all backgrounds and perspectives beyond just gender to join our community. At eResearch Australasia 2024, we hosted a Birds-of-a-Feather (BoF) that showcased community members and their career paths across a range of roles, from technical staff through to leadership roles. It was very successful, highlighting both the challenges and opportunities to advance diversity and inclusion in our sector. We want to continue that positive discussion at eResearch NZ / eRangahau Aotearoa. This BoF will showcase the diversity of our HPC and eResearch communities and create a forum for learning from others' experiences and perspectives.

To drive our discussion, we'll start with a panel of 3-4 invited speakers who represent the breadth of the community – diverse in role, domain, and career stage. We'll begin with lightning talks by the panel speakers, then move into a moderated discussion and open the floor for general questions and discussion.

Lightning talk speakers:
- Dr Rina Hannaford, Senior Data Scientist, AgResearch
- Dr Kari L. Jordan, Executive Director, The Carpentries
- Dr Jana Wold, Post Doctoral Fellow, University of Canterbury

ABOUT THE AUTHORS
The Australasian Chapter of the global organisation Women in High Performance Computing (WHPC) was launched in 2019. The founding organisations are New Zealand eScience Infrastructure (NeSI), NCI Australia, Pawsey Supercomputing Research Centre, Monash University, and Australasian eResearch Organisations (AeRO). The Chapter supports an active Slack community, regular BoFs at conferences, online meetups, and practical resources for organisations to create positive change. We welcome all backgrounds and perspectives beyond gender to join our Chapter.

The Australasian Chapter Organising Committee is currently composed of:
- Jana Makar, New Zealand eScience Infrastructure (NeSI)
- Aditi Subramanya, Pawsey Supercomputing Research Centre
- Emily Barker, University of Western Australia
- Linda McIver, Australian Data Science Education Institute
- Carina Kemp, Amazon Web Services (AWS)
- Kiowa Scott-Hurley, Independent
- Kerri Wait, Independent
- Loretta Davis, Independent

For more information, visit https://tinyurl.com/whpcaunz.

Challenges of Secure File Transfer in Collaborative Research

Yeshaswini (Yesh) Ramesh,
REANNZ
yeshaswini.ramesh@reannz.co.nz

ABSTRACT
To support effective collaboration, researchers require efficient and secure methods to share data especially large data sets. However, when sharing sensitive or confidential data, security is critical, requiring encryption and careful management of permissions to protect data both in transit and at rest. Additionally, it's essential to ensure that data reaches the right person.
To address some of these challenges, the NREN community has come together to develop and leverage purpose built solutions for the academic research sector using open source software solutions such as FileSender and Globus.

1. Filesender overview:

FileSender is a powerful and user-friendly file-sharing service designed to simplify the process of exchanging files securely and efficiently over the internet. In New Zealand, Filesender is provided by REANNZ and hosted in the country. The advantages of Filesender service include:

- Ability to move data efficiently, securely, and reliably between users within New Zealand or around the world.
- Send and receive files via your web browser, without extra software or plug-ins.
- Designed for scale: Send multiple large files and transfer up to 30 files at a time with a combined maximum size of 5 TB per upload.
- Senders have full control over who receives and can access the files and for how long.
- Security and privacy are guaranteed; we treat your files with the utmost confidentiality.
- With end-to-end encryption, it is technically impossible to access files 'from outside'.

Use case UMC Amsterdam:
FileSender has proven to be an essential tool for UMC Amsterdam, facilitating secure and efficient data sharing among researchers. This case study underscores how FileSender has enabled the transfer of large datasets, including medical imaging files and genomic data, which are critical for collaborative research projects. By utilizing FileSender, UMC Amsterdam ensures the encryption and secure management of sensitive patient data, both during transit and at rest. This has considerably enhanced the efficiency of data sharing processes, allowing researchers to concentrate more on their work rather than on the logistics of data transfer.

2. Globus overview:

Globus is designed for the needs of the academic research community and is supported by most by most HPCs in the world and provides a more efficient way to transfer large files to known destination points from kilobytes to petabytes. In New Zealand, Globus is supported by NeSI and REANNZ.

Globus allows researchers to access their research data from local storage, institutional storage, tape archives, commercial cloud storage, HPC systems and scientific instruments all from a single web browser. Globus also variety of features with data management and sharing capability including:

- Move data efficiently, securely and reliably between rooms, across c
  around the world.

- Visibility of data across different institutions or supercomputer facilities using campus credentials.
- Share files with colleagues with only their email address, without needing to set up a new user account.
- Data can be secured with multi-factor authentication and federated login, with full control over data access and sharing permissions of users.
- Use the open REST APIs, Python SDK or internal workflow engine to monitor and access transfers, tasks and remote directories, or build custom research data services and applications.

Use case: **The Research Computing Centre at the University of Queensland (UQ)**
AARNET and Globus have rolled out Globus at the Research Computing Centre at the University of Queensland (UQ). The University has noted the benefits of this tool, highlighting how it enables secure and fast data sharing, making collaborations easier and faster. Before the introduction of Globus at UQ, researchers had to courier physical hard drives containing research data, which was both expensive and risky. The integration of Globus into UQ's MeDiCI data fabric now allows researchers to move data seamlessly into and out of existing UQ collections.
This presentation will discuss the challenges surrounding file transfer within a collaborative research environment, explore in more detail how the NREN community is developing solutions to support the academic research community and provide insight into some additional use cases from around the globe.

ABOUT THE AUTHOR
Yeshaswini (Yesh) Ramesh is Head of Network Operations at REANNZ.
She brings her extensive experience as a Senior Network Engineer both at REANNZ and earlier on her career worked for the IBM NZ team supporting the Vodafone IP Network. In India, she worked as senior engineer at Juniper Networks, a Business Analyst at Capgemini and had various engineering roles at Alcatel-Lucent where she started her career.
Since joining REANNZ in 2015, Yesh has worked as a part of the network operations team that implement solutions, new products, and support projects for REANNZ members. Yesh builds and maintains relationships with members of the network, vendor's technical teams and other National and Research Networks (NRENs) globally.

Changes to digital identity concepts and the implications for eResearch

David Brownlie
REANNZ
david.brownlie@reannz.co.nz

ABSTRACT

Digital identity is gaining momentum across the planet with some estimates suggesting that 6.5 billion people worldwide will be using a digital identity by the end of 2026. The rapid adoption of digital identity has had the flow on effect of both increased investment in, and increased regulation of, digital identity around the globe. As a result of these trends digital identity concepts are changing to maximise trust and security whilst maintaining utility.

These changes will impact researchers and eResearch infrastructure access sooner rather than later. However, they also present many opportunities to more easily support collaborative projects, incorporate te ao Māori concepts of identity, make effective use of artificial intelligence or machine learning and protect intellectual property, whilst simultaneously reducing administrative overheads associated with managing access to data, infrastructure and services.

This presentation will cover emerging trends such as the decentralisation of identity, passwordless verification, machine identity, regulation in Aotearoa and what this could mean for researchers and eResearch infrastructure.

ABOUT THE AUTHOR

David Brownlie has been working in technology and telecommunications for over 30 years and has been with REANNZ since 2006 working in various technology based roles. He is currently part of the REANNZ engagement team fulfilling the role of Senior Technical Advisor.

Containerisation to enable binary portability – A use case from climate research

Alexander Pletzer[1], Andrew Pauling[2], Chris Scott[3], Murray Cadzow[2]
1 NeSI/NIWA, 2 University of Otago, 3 NeSI/University of Auckland
alexander.pletzer@nesi.org.nz

ABSTRACT

Climate change has far reaching implications, affecting many sectors of society and the economy. Climate research rests on running large and intricate numerical models to simulate the effect of various emissions scenarios on temperature and other meteorological variables. The Unified Model (UM) used by Dr Pauling is one of such codes.
Compiling the UM can be a daunting task, as the code not only relies on thousands of files but also configuration files, which point to data stored on remote sites, custom tools to fetch data and code, and a custom build system that determines the order of compilation of Fortran files.
With the advent of new NeSI computing in 2025 to replace the capability supercomputer Maui, NeSI has devoted resources to help scientists migrate to our AMD Milan based cluster. To facilitate the transition, NeSI has built an Apptainer container with a version of UM, which can be ported to any platform running on a recent Linux operating system. This means that researchers can share their executables with colleagues and/or run on the platform of their choice, whether on NeSI or on another supercomputing site. Here we present our experience in containerising the UM with an emphasis on lessons learned and gotchas.

ABOUT THE AUTHOR

Alex Pletzer helps scientists run better and faster on NeSI platforms. Alex has a PhD in physics. But over the years, he found himself drifting towards computing and considers himself now a research software engineer. You can find him windsurfing in Wellington and watercolour painting in his free time.

Continuing the research data count and what's next?

Ai-Lin Soo, Monash University, ai-lin.soo@monash.edu
Rhys Francis, Research Data Culture Conversation, rhyssfrancis@gmail.com
Luc Betbeder-Matibet, UNSW, luc@unsw.edu.au
Claire Rye, New Zealand eScience Infrastructure, claire.rye@nesi.org.nz
Nick Jones, New Zealand eScience Infrastructure, nick.jones@nesi.org.nz

ABSTRACT

Aim: to continue the research data conversation that has evolved over the last few years at eResearch NZ and at eResearch Australasia.
Working with our Research Data Culture Conversation (RDCC) colleagues from Australia, we have been asking a series of apparently simple questions to the Aotearoa New Zealand research data community. This work has produced the first 'Macro View' of our Research Data resulting in both an estimate of the total research data holdings in Aotearoa New Zealand - 45 PB (Petabytes) at the end of 2022, and a series of observations, such as the difficulty of the exercise and that measuring and valuing of research data is mostly absent. The insights gained demonstrate that without sustained work effort and established processes, repeatedly measuring our data holdings will be difficult. Recognising this, the Council of Australasian University Directors of Information Technology (CAUDIT), have initiated the first year of benchmarking its members' data holdings, focussing on data related to education, business and research. The conversation this BoF continues is grounded in measuring. With this survey, and with the help of industry partners, we look to improve Aotearoa New Zealand's data holding assessment. We will update the community on measuring with hopefully some 2024 numbers.
The difficulty of answering the simple question, how much unique research data does New Zealand have, indicates that even though Research Data is considered a valuable asset by Researchers and their Institutions, the institutions know very little about the data they are holding.
Our discussion of key questions continues:

- What do we know about our research data as a national research enabling asset?
- Is everything we are holding actually data, let alone valuable?
- What fraction of our 'data asset' is in fact the digital debris of research?
- What do we need to know?

ABOUT THE AUTHORS

Ai-Lin Soo is currently a Project Manager at Research Technology Services, UNSW Sydney. She is involved in a number of cross cutting initiatives, applying her project management skills to a diverse range of projects predominantly focussed on research data management. Ai-Lin is also heavily involved in the Research Data Culture Conversation (www.researchdataculture.org). Ai-Lin has a background in Commerce and BioMedical Science and is also the Operations Manager for the Ai for Law Enforcement and Community Safety Lab at Monash University.

Claire Rye is a Product Manager at New Zealand eScience Infrastructure (NeSI) based out of the University of Auckland. She is responsible for the National Data Transfer Service and works across the Aotearoa Genomics Data Repository and Rakeiora Pathfinder projects and looking at research data management and data lifecycle more generally across NeSI. Claire holds a PhD in organic chemistry and has spent the last 11 years working in the UK in a variety of research settings.

Rhys Francis has contributed to Australian eResearch through many activities including developing the initial investment plan in eResearch for the National Collaborative Research Infrastructure Strategy; proposing the eResearch investments in the scale up created by the Super Science Initiative and serving as the Executive Director of the Australian eResearch Infrastructure Council for seven years. Following on he has developed a revised eResearch Framework for government, assisted the University of Melbourne develop a Petascale Campus Initiative and the Defence Science and Technology Group review eResearch infrastructures. He currently works with the University of Auckland on their eResearch strategy, facilitates the Research Data Culture Conversation (www.researchdataculture.org) and is the Strategic Advisor, Australian BioCommons (www.biocommons.org.au).

Luc Betbeder-Matibet is a nationally recognised subject matter expert in eResearch, University Research Data Management and shared computational infrastructure services for researchers. He has held director-level roles for 15 years in ICT and eResearch. Luc is the Director Research Technology Services at UNSW, a shared services function that he established which is responsible for Research Computing and Research Data. He is an Adjunct in UNSW Faculty of Medicine Centre for Big Data and has been a Visiting Scientist with the Visual Analytics Team in CSIRO Data61. Recently Luc has been working with colleagues to count how much Research Data there actually is in Australia.

Nick Jones is a leader of the national eScience and research High Performance Computing platform investment in New Zealand, NeSI - National eScience Infrastructure. He has been in leadership roles in the academic research sector for 20 years across eLearning and eResearch. Nick led the development of the eResearch Ecosystem Framework and sponsored a subsequent International Benchmarking Study. Nick has been an advocate for the evolution of research practice in advanced digital technologies across research data and software, computational thinking, and a user-centred and relational approach to delivering innovative experiences to researchers.

Deep diving into deep learning and AI: experiences and lessons learned

Brent Martin
Manaaki Whenua – Landcare Research
martinb@landcareresearch.co.nz

ABSTRACT

Around a decade ago Deep Learning emerged as the new frontier of computer vision, eclipsing the performance of previous methods for image classification, semantic segmentation and object detection/segmentation. At Manaaki Whenua Landcare Research we have successfully trialled deep learning for all three of these uses and applied it to many domains, including detecting and monitoring invasive predators (classification), mapping land cover, land use, wetlands and wilding conifers (semantic segmentation), and mapping individual trees (object segmentation). More recently we have "stretched" the use of deep learning still further by using it automate the generation of landscape-scale polygons for soil mapping and other uses, by teaching the model to perform this previously manual task.

While investigating deep learning's use in these domains, we have also been developing inhouse tools, including user-friendly software for running and fine-tuning image detection models (CamTrapNZ), and for applying semantic and object segmentation to remote sensing imagery (DeepSeg). While these tools have been extremely valuable, they are still relatively immature. Further, the expertise required to use them effectively is largely contained within a small pool of specialised data science/remote sensing researchers. To overcome this, we are moving to a new stage where we are maturing the tools to make them more accessible to the wider research community, as well as developing online training material to enable uptake by non-data scientists. A key challenge for this effort is the wide range of abilities and expectations within the research community (our "customers"), requiring a range of solutions. Adding to this is the constant evolution of deep learning itself, meaning both educational material and tools quickly become obsolete.

More recently, Manaaki Whenua has embraced Generative AI as a highly promising addition to the toolset for a range of applications including information extraction, scenario visualisation and simulation. Unlike deep learning, which had obvious application to an existing set of problems and was therefore adopted in a fairly systematic manner, the use of GenAI is more general and speculative, and is therefore emerging organically throughout the organisation. Key to the successful adoption of GenAI is the sharing of knowledge and experiences across an extremely diverse range of projects.
In this presentation we will share our experiences over the past five years, as well as our near-term plans for maturing our tools, upskilling Manaaki Whenua's researcher base and sharing project models, data, software and knowledge.

ABOUT THE AUTHOR

Brent Martin is a senior data scientist and machine learning specialist at Manaaki Whenua – Landcare Research. His 38-year career spans both academic research as a senior lecturer at Canterbury University, as well as software engineering and R&D roles in various commercial companies from local software house Jade to Google NY.

Brent's research in AI and machine learning includes developing new ML classification algorithms, applying ML to real-world problems such as electricity demand forecasting and internet search engines, research and development in Intelligent Tutoring Systems, developing social network analysis and anomaly detection techniques for criminal investigation, and applying deep learning to environmental computer vision proble

At Manaaki Whenua, Brent has applied deep learning to problems for a range of clients including government departments such as MfE and MBIE, and regional councils throughout New Zealand, as well as internal projects aimed at automating and upscaling some of the services Manaaki Whenua already provides. More recently this research has included applied generative AI. Brent holds a PhD in Computer Science (artificial intelligence) from the University of Canterbury, New Zealand.

Driving Digital Research in Australasia with the Nectar Research Cloud

Bernard Meade
Australian Research Data Commons
bernard.meade@ardc.edu.au

ABSTRACT

Established in 2012 as part of the Australian Research Data Commons (ARDC), the Nectar Research Cloud is a national research cloud infrastructure operating across multiple research institutions in Australasia. This distributed model has evolved into a robust platform supporting over 4,000 researchers with flexible, secure, and powerful cloud computing services. As research demands grow and technology advances, Nectar continues to adapt and evolve, supporting both traditional and emerging fields, including data-intensive disciplines such as artificial intelligence and climate science.

Through strategic partnerships and enhanced GPU computing resources, Nectar empowers researchers to tackle increasingly complex problems, expanding opportunities for innovation and collaboration across disciplines and borders. Our services—such as Virtual Desktops and interactive platforms like JupyterHub—facilitate accessible and reproducible research practices, ensuring that researchers across institutions can fully engage with data-driven inquiry and discovery.

Looking to the future, Nectar is expanding geographically and enhancing its resilience through new nodes, with a particular focus on cybersecurity and service quality. Our new National Node, developed in collaboration with AARNet, will strengthen Nectar's core services and bolster our cybersecurity strategy as we work towards ISO27001 certification to ensure the highest standards of security and service management. The introduction of a High Throughput Computing service will further broaden our capacity to support extensive data processing needs, reinforcing Nectar's position as a cutting-edge eResearch infrastructure that strengthens research capabilities in Australasia.

In this presentation, we will demonstrate how Nectar’s evolving portfolio of services embodies the theme of **"Evolution of eResearch Infrastructure"** by meeting the current and future needs of higher education and research across the region. We will highlight how Nectar's continual adaptation positions the research community in Australasia to address complex challenges and contribute meaningfully to global knowledge. By fostering a dynamic and resilient infrastructure, Nectar continues to empower researchers to achieve excellence, foster collaboration, and drive impactful, sustainable advancements.

ABOUT THE AUTHOR

Bernard Meade is the Manager of ARDC Nectar Research Cloud operations, bringing over 28 years of IT experience in Higher Education. With a PhD in astronomy instrumentation, he has dedicated two decades to supporting research computing infrastructure. Since the Nectar Research Cloud's launch in 2012, he has played a key role in its operations, combining technical expertise with a deep understanding of researchers' needs.

Embracing a changing eResearch landscape: NeSI's horizons beyond traditional HPC

Jun Huh, NeSI, jun.huh@nesi.org.nz
Claire Rye, NeSI, claire.rye@nesi.org.nz
Blair Bethwaite, NeSI, blair.bethwaite@nesi.org.nz

ABSTRACT

In 2024 NeSI was fully focused on provisioning its new HPC infrastructure. This marks the third major infrastructure provisioning process NeSI has embarked on. This presentation focuses on two of the key approaches taken and shares the experience and learnings from the process.

The first approach is around provisioning of the HPC and its core services using the Infrastructure-as-Code (IaC) approach. NeSI has an ongoing collaboration with AgResearch to build and operate an eResearch infrastructure. This was done on top of NeSI's Flexible HPC, a platform designed to support flexibility and scale. Many of the learnings from this collaboration was applied to NeSI's own infrastructure provisioning process, and much of the IaC based approach helped NeSI in taking the processes from the code base and modifying them to fit the needs. Through this journey, NeSI has matured greatly in its approach in code iteration and deployment pipelines.

The second approach is around NeSI's data migration into the new infrastructure. NeSI has adopted new technologies, WEKA and Versity, to address the needs around ever growing data. WEKA is used for providing tiered active storage, and Versity for NeSI's new Freezer service, offering a long term archive solution. NeSI has applied thinking of data in temperature to allow smoother migration as well as ongoing HPC operations. This presentation will address many of the challenges around data migration and the solutions applied.

ABOUT THE AUTHORS

Jun Huh is a Product Manager at NeSI. Jun brings his experience from start-up industries into the field of eResearch. He has been involved in genomic data management related projects for the past 2-3 years, to help build a data repository system in partnership with Genomics Aotearoa, and more recently, prototyping for Rakeiora Pathfinder project, which focuses on enabling research while retaining full visibility and control of data to the indigenous communities.

Claire Rye is a Product Manager at New Zealand eScience Infrastructure (NeSI) based out of the University of Auckland. She is responsible for the National Data Transfer Service and works across the Aotearoa Genomics Data Repository and Rakeiora Pathfinder projects and looking at research data management and data lifecycle more generally across NeSI, and is a co-chair for the World Data System Early Career Researchers network. Claire holds a PhD in organic chemistry and has spent the last 11 years working in the UK in a variety of research settings.

Blair Bethwaite is NeSI's Solutions Manager, when not on a bike, squash court or in the ocean, he specialises in innovation of the underlying systems and platforms to enable NeSI's teams to deliver greater value to researchers. He has almost 20 years of experience in distributed computing; both in research and for research; for institutional and national projects; from science application collaborations, through grid & cloud middleware development, to full HPC & cloud system infrastructure design, implementation, and operations.

eResearch in an Age of Austerity

Wes Harrell
Manaaki Whenua
HarrellW@landcareresearch.co.nz

ABSTRACT

Proposal: Lightning Talk in support of a BOF: Building a Skilled Research Workforce

Issues:

- How can organisations be effective with limited resources
- How do we support a geographically dispersed workforce?
- CRI staff are billable - how can we "pay" for training?
- CRIs are relatively small with limited turnover - how can we efficiently upskill our team?

Proposition:
We must collaborate, co-ordinate and communicate:

- NeSI as a conduit for co-ordination and communication
- Regionally focused training - collaborating with other nearby organisations (CRIs, Unis, Govt departments) to host training events.
- Implement the most effective communication tools that foster better engagement, teaching and consulting

ABOUT THE AUTHOR

Wes Harrell is eResearch Support Manager at Manaaki Whenua. Wes and the rest of the eResearch team assist with software engineering, data engineering, HPC, devops, training and other things digital.
Manaaki Whenua is the Crown Research Insitute focused on our land and environment.

Evolution of Research Infrastructure at University Technology Sydney: Harnessing the Power of Digital Science Solutions

Maxine Bryant
Digital Science
m.bryant@digital-science.com

ABSTRACT

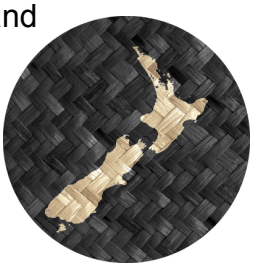
The University of Technology Sydney (UTS) has embarked on an evolutionary journey to enhance its research infrastructure, partnering closely with Digital Science to create a seamless, integrated environment for managing and showcasing research. This abstract delves into the transformative impact of Digital Science products at UTS, outlining the phased adoption and integration of various tools that together support data-informed decision-making, streamline the research lifecycle, and amplify the visibility of research outputs.

Since 2013, UTS has viewed Digital Science as a strategic partner, fostering a collaborative relationship that has evolved with the growing demands of research management. Initially, UTS selected Symplectic Elements as its core system for managing research outputs and profiles, marking the beginning of a robust research information management ecosystem. This selection was pivotal in supporting academics and professional staff in recording a wide array of scholarly activities, including research outputs, engagement activities, and teaching contributions. With Elements, UTS laid the groundwork for centralized, accessible research data, enabling profiles to populate across the UTS website and UTS Research Insights dashboards, and allowing streamlined data collection through the annual Research Outputs Collection (ROC).

Building on the success of Symplectic Elements, UTS progressively introduced additional Digital Science products. In 2015, the Altmetric Explorer was integrated to provide nuanced insights into the reach and impact of UTS research, helping to track engagement across various digital platforms. The implementation of Dimensions in 2017 further enriched the university's analytical capabilities, allowing for comprehensive bibliometric analyses that informed research strategy. The launch of UTS public profiles through Elements in 2019 (https://profiles.uts.edu.au/) enabled the institution to amplify researchers' work and connect with a broader audience, including potential collaborators and funding bodies. Most recently, UTS integrated Figshare in 2024 to enhance its repository capabilities and leverage DOI minting through the Australian Research Data Commons, expanding discoverability for non-traditional research outputs.

Through this phased approach, UTS has built a cohesive research infrastructure that enhances both administrative efficiency and research visibility. The use of Symplectic Elements for ROC and public profiles has improved the accuracy and accessibility of research data, while Altmetric Explorer and Dimensions have empowered UTS with metrics-driven insights to guide research strategy and assessment. The recent addition of Figshare is set to further drive impact, particularly for non-traditional research outputs, by expanding the digital footprint of UTS's research portfolio. UTS's partnership with Digital Science exemplifies a sustainable, scalable model for institutions aiming to adapt to a rapidly evolving research landscape.

"Digital Science has supported UTS in our data informed decision-making, storytelling and streamlining the research lifecycle in an increasingly complex research environment. Our relationship with Digital Science goes beyond this as one of our most receptive and trustworthy partners." – Sofia Haidar, Deputy Director, UTS Research Office



ABOUT THE AUTHOR

Maxine Bryant is passionate about helping people adopt new technologies, simplify processes, and create positive cultural change to drive better business outcomes. Experienced in leadership and governance in the tertiary and research sectors, Max works as Solutions Consultant for Digital Science, specializing in the implementation of Symplectic Elements as a research information management solution.

Expanding Cluster Access: Integrating Open OnDemand with Otago's Research and Teaching

Peter Higbee and Murray Cadzow
University of Otago
peter.higbee@otago.ac.nz, murray.cadzow@otago.ac.nz

ABSTRACT

Cluster computing and integrated research workflows are essential for modern research, enabling high-throughput data analysis and complex computational tasks. Open OnDemand provides a web-based interface for accessing and managing high-performance computing (HPC) resources, significantly simplifying the user experience for researchers across various disciplines. This platform removes the need for extensive technical knowledge, as it offers intuitive interfaces for job management, data transfer, and interactive applications, including Jupyter Notebooks, RStudio, and remote desktops.

At the University of Otago, we have undertaken an initiative to expand access to computational resources by integrating Open OnDemand with existing research compute infrastructure, alongside the deployment of custom applications that align with research and teaching needs. This initiative allows researchers seamless access to compute resources while also providing students with practical exposure to these systems in classroom environments.

This talk will showcase the current design and implementation of the Open OnDemand platform and give examples of how it has been utilised for both research and teaching activities at University of Otago.

ABOUT THE AUTHORS

Dr Murray Cadzow is a Scientific Programmer within Research Teaching IT Support at the University of Otago, with a background in bioinformatics. Murray has been heavily involved in computational literacy and bioinformatic training at the University of Otago - co-organising Research Bazaar and the Otago Bioinformatics Spring School. He is certified as both a Carpentries instructor, and instructor trainer.

Peter Higbee is a Systems Engineer at the University of Otago with a background in Earth Science. He manages HPC systems, storage solutions, and applications across campus, bringing over 20 years of expertise in IT. In collaboration with Otago's Research and Teaching leadership and engineering team, Peter has worked on numerous initiatives to streamline and advance research infrastructure, including the recent integration of Open OnDemand to support both research and educational applications.

Exploring NeSI's Cloud-Based OpenOnDemand Training Environment: A Quick Demo

Kahu Anderson & Nisha Ghatak
NeSI

Situation:

The complexity of accessing and utilising high-performance computing (HPC) infrastructure can create a significant barrier for researchers and students. Recognising this challenge, NeSI has introduced a cloud-based OpenOnDemand training environment, designed to simplify and streamline access while fostering hands-on skill development in computational research.

Task:

To introduce the capabilities of NeSI's cloud-based OpenOnDemand platform through a concise demo, highlighting its potential to empower researchers with intuitive tools for HPC usage, training, and upskilling. The goal is to demonstrate how this environment supports equitable and efficient access to advanced computing resources.

Action:

This lightning talk will showcase a live demonstration of NeSI's OpenOnDemand environment, focusing on key features such as the seamless user interface, remote file management, job scheduling, and access to interactive applications like Jupyter and RStudio. It will also touch on the cloud-based deployment, which enhances accessibility and scalability for training programs. The session will briefly highlight how this environment can be customised for workshop scenarios to cater to diverse research and learning needs.

Result:

Attendees will gain a practical understanding of NeSI's cloud-based OpenOnDemand training environment and its role in simplifying HPC usage. The session will demonstrate how this tool fosters learning, reduces barriers to offering digital training, and equips researchers with a robust platform to drive computational advancements in Aotearoa's research landscape.

ABOUT THE AUTHORS

Kahu Anderson is a Research DevOps Support Engineer at NeSI with a strong background in Information Technology. With NeSI's new Research Developer Cloud, he provides the Researcher Support team with information and documentation to facilitate collaboration between researchers and NeSI's new platform.

Nisha Ghatak is a Research Communities Advisor – Training Lead at New Zealand eScience Infrastructure. She facilitates skill development efforts nationally and supports training communities through her role as a community coordinator for The Carpentries in Aotearoa New Zealand. Nisha is also a certified Carpentries instructor and part of the Board of Directors for The Carpentries.

Flexin for three years.
An update on AgResearch's eResearch Infrastructure refresh with FlexiHPC.

Ben Taylor
AgResearch
Ben.Taylor@agresearch.co.nz

ABSTRACT

On the 9th of July 2021 AgResearch and NeSI signed a Memorandum of Understanding to work together as partners on areas of shared benefit, thus kickstarting our eResearch infrastructure refresh journey to replace our legacy HPC and storage environments with fit for purpose infrastructure delivered via HPE and NESI's FlexiHPC platform.

Our task was ambitious, wanting a single platform for all AgResearch's scientific data and compute needs beyond a traditional HPC cluster. Including proprietary windows compute environments and the amalgamation of several filesystems into one.
Fast forward to 2025, this whirlwind talk will cover off our journey as we navigated covid lockdowns, supply chain disruptions, countless sprint ceremonies, and data migrations. To deliver a platform that delivers value for our scientists. As well as the ups, downs, and lessons we learned along the way.

ABOUT THE AUTHOR

Ben Taylor is the eResearch, Infrastructure and Security Manager at AgResearch with over 20 years of experience. By day (and sometimes by night), he juggles a range of responsibilities to ensure AgResearch's researchers have the infrastructure and services they need to feed their ever-growing data appetite. For his sins, he also tries to balance this with the need to protect information and data. Easy right?

Following the yellow brick road - our Research Data services journey

Yvette Wharton[1], Laura Armstrong[1], Marcus Gustafsson[1] and Mark Gahegan[1]
[1] Waipapa Taumata Rau, University of Auckland
y.wharton@auckland.ac.nz, l.armstrong@auckland.ac.nz, m.gustafsson@auckland.ac.nz, m.gahegan@auckland.ac.nz

ABSTRACT

With increasing pressure on storage needs, legal and ethical considerations, technological advancements, and a push for open research, the last 10 years have seen the evolution and expansion of Research data services across the research data lifecycle. This evolution has seen a need for more integration and collaboration among service providers to enable effective data management planning, collection, storage, and analysis, as well as the publication of data and research artefacts.

Just like Dorothy on the yellow brick road, we are on a journey of discovery with our friends with research data and storage. Like Dorothy, we begin our journey in Kansas before being whisked away by a tornado to a new land. This presentation will discuss our adventure, companions, challenges, emerald city goals, growth, and learning.

ABOUT THE AUTHORS

Yvette Wharton works at the Centre for eResearch, Waipapa Taumata Rau, University of Auckland. She is the Technical Business Lead for the Research Data Management Programme, working on the Secure Research Environment and machine-actionable Data Management Planning initiatives. http://orcid.org/0000-0002-6689-8840

Laura Armstrong is the eResearch Engagement Lead for Waipapa Taumata Rau, University of Auckland. She collaborates to engage with the research community to raise awareness and use of modern technologies and tools to advance research. Her focus in recent years has been maturing research data management through researcher development, service delivery and policy. http://orcid.org/0000-0003-2370-3924

Marcus Gustafsson is the eResearch Operations Manager at the University of Auckland. He hails from a physics and computation background but has moved through HPC infrastructure service delivery to enabling research IT services and researcher upskilling with teams of eResearch Specialists. In his current role he is mostly a self-professed black-hat and practicing disciplinarian trying his utmost to keep up with the ever-growing business case writing and compliance rat-race.

Mark Gahegan is the founding director of the Centre for e-Research and Professor in the School of Computer Science at the University of Auckland, New Zealand. Mark has wide interests at the intersection of Computer Science, Machine Learning, eScience and GIS Science. This is because he can never sit still for too long and is easily distracted. He is kind of scruffy-looking. He also directs the ‘Beyond Prediction...’ national Data Science programme, focused on applying emerging Machine Learning, eScience and GenAI methods to problems in Ecology, genomics and infectious diseases.

From Cloud to HPC and back again – migration, rebasing and upcycling of NeSI's platforms

Blair Bethwaite
New Zealand eScience Infrastructure, University of Auckland
blair.bethwaite@nesi.org.nz

ABSTRACT

Over the last 12 months we at NeSI have been busy Rebasing, but what does this mean and what is happening down in the guts of the systems that provide the computational power behind researchers? Join me for a whirlwind tour of the technologies, datacentre jenga, and pie-in-the-sky dreams behind the infrastructure underpinning NeSI's current platform refresh and migration. You won't hear Nick Jones mention a CPU model or network stack, but I might.

In this talk I'll discuss some of the goals and motivations behind NeSI's infrastructure refresh, cover it's high-level design, and how it marries Cloud and HPC to create new eResearch solution opportunities for New Zealand's RSEs and researchers. Are we bringing HPC to the Cloud, or turning HPC into Cloud, maybe it's both – come along to find out.

ABOUT THE AUTHOR

Blair Bethwaite is NeSI's Solutions Manager. When not on a bike, squash court, or in the ocean, he specialises in innovation of the underlying systems and platforms to enable NeSI's teams to deliver greater value to researchers. He has almost 20 years of experience in distributed computing; both in research and for research; for institutional and national projects; from science application collaborations, through grid & cloud middleware development, to full HPC & cloud system infrastructure design, implementation, and operations.

From Static to Dynamic: LivePublication and the Quest for Reproducible, Living Articles

Augustus Ellerm[1,2], Ben Adams[2], Mark Gahegan[1], Nelis Drost[1]
[1]University of Auckland, [2]University of Canterbury
m.gahegan@auckland.ac.nz, gus.ellerm@pg.canterbury.ac.nz,
benjamin.adams@canterbury.ac.nz, nelis.drost@auckland.ac.nz

ABSTRACT

Within the computational sciences, it is intuitive that our research publications are primed to derive more value from integrations across eScience technologies, with substantive efforts being made towards supporting reproducibility and reuse. Early Research Software Environments (RSEs) and Executable Articles (EAs) paved the for how we conceptualise computational tools and their relationships to our publication containers. More modern platforms such as WholeTale and Binder Hub containerise Literate Programming Environments (LPEs), combining reproducible code, data, and narrative, which have been popularised in the academic publication context.

In this talk I expand on the previously discussed LivePublication framework---a framework designed to enable the automated generation and modification of scientific articles. LivePublication addresses the challenges of reproducibility within academic articles, and develops upon existing technologies to enable live and dynamic updates to academic articles as new results are generated from experiment interfaces. LivePublication does not set out to create a bespoke, centralised platform for the deployment of live, reproducible publication containers. Instead, it seeks to integrate existing off-the-shelf eScience technologies, from Workflow Management Platforms (WMPs) to LPEs.

During the last year, with collaboration from Globus Labs (https://labs.globus.org) and Argonne National Laboratory (https://www.anl.gov), we focused on developments in computational workflow provenance, and how we may maintain truth across multiple runs of a computational experiment. This talk will address 2 questions:

1. How can we record granular provenance across distributed, heterogeneous resources for application in **live** articles? This will be addressed with a standard provenance model for LivePublication, with an example application in the WMS Globus Flows.
2. What challenges, and opportunities, were associated with approaching LivePublication as an integration of maturing eScience technologies, and how may we focus on designing tools that support novel implementations? This will be addressed as a general discussion of my experience working with open technologies, communities, and standards to meet our goals.

I will provide an outline of the LivePublication Provenance model, and application of the provenance model to a production WMS (Globus Flows, https://docs.globus.org/api/flows/).

ABOUT THE AUTHORS

Mark Gahegan is Professor in Computer Science at the University of Auckland, where he also directsthe Centre for eResearch. He is PI of '*Beyond Prediction…'*, a large, 7-year Data Science Programme Grant from MBIE. His research interests are in eScience, GIScience, Data Science and all points in between.

Gus Ellerm is a PhD student in Computer Science at the University of Canterbury and research fellow at the University of Auckland. Currently, they are studying research workflows and their role in supporting *live* publications. Gus leads the implementation of the work reported here, is funded via the above MBIE grant and is supervised by Ben Adams and Mark Gahegan. Gus has recently been invited to present his work at the Globus World`24 conference.

Ben Adams is Associate Professor of Computer Science and Software Engineering at the University of Canterbury. His research interests revolve around new ways to use computing technology to help advance human understanding of our environment and world, drawing from data science, spatial science and cognitive science.

Nelis Drost Cornelis is a complex systems modeller with experience in fields including ecology, archaeology, oncology and epidemiology. Having experience developing both models and analyses in an academic setting, and production software in a commercial setting, he now works as a Snr. Solutions Specialist at the Center for eResearch, where he supports researchers with modelling, data visualization and software development. https://orcid.org/0000-0002-7355-9978

Globus Compute: Federated FaaS for Integrated Research Solutions

Ryan Chard, Yadu Babuji, Josh Bryan, Rachana Ananthakrishnan, Kyle Chard, and Ian Foster
Argonne National Laboratory and University of Chicago
rchard@anl.gov

ABSTRACT

The ever-increasing amount of data produced by scientific experiments and simulations makes it essential to employ HPC resources to extract meaningful insights from the raw data. HPC enables researchers to perform simulations, modeling, and analysis, which are critical to predicting outcomes, guiding experiments, and developing new technologies [1]. However, the pervasive use of HPC presents new challenges as researchers must familiarize themselves with complex environments, batch schedulers, and authentication schemes. Globus Compute [2] is a Function-as-a-Service platform designed to provide a scalable, secure, and simple interface to HPC resources. Globus Compute implements a federated model via which users may deploy endpoints on arbitrary remote computers, from the edge to high performance computing (HPC) cluster, and they may then invoke Python functions on those endpoints via a reliable cloud-hosted service. Unlike traditional cloud-hosted FaaS platforms (e.g., Amazon Lambda) and open-source platforms (e.g., OpenWhisk) [3], Globus Compute combines a single cloud-hosted web service with an ecosystem of edge endpoints. Recently, Globus Compute released a multi-user endpoint, designed to be deployed and managed by system administrators. These endpoints secure access to resources and provide fine-grained auditability, while trivializing access for users.
In this talk, we will discuss how Globus Compute can be used to simplify and federate access to heterogeneous HPC resources. We will discuss the design, implementation, and deployment of Globus Compute and discuss how multi-user endpoints can be used by administrators to securely manage system access. Additionally, we will showcase how Compute can be used to enable integrated research solutions that span institutions.

ABOUT THE AUTHOR

Ryan Chard joined Argonne National Laboratory in 2016 as a Maria Goeppert Mayer Fellow. He is currently a Consultant at Argonne focusing on the development of cyberinfrastructure to accelerate scientific research. In particular, he works on tools to streamline data analysis pipelines and develops integrated research solutions that incorporate HPC resources into scientific processes. He has a Ph.D. in computer science and an M.S. from Victoria University of Wellington, New Zealand.

Grass roots data management support at Plant & Food Research

Briget Lander
Plant & Food Research NZ Ltd
Briget.Lander@plantandfood.co.nz

ABSTRACT

Plant & Food Research (PFR) is a CRI focusing on improving the way we grow, fish, harvest, prepare and share food. We're based across 14 locations in NZ and employ around 700 science staff. Spread out across all our science teams is a network of Data Stewards whose role is to help and advise their team mates on how to improve their own data management practice. The network was set up six years ago and Data Stewards are now an integral part of how we do science at PFR. The network is well supported at all levels of the organisation because it brings value in many different ways.
This presentation is about why and how the network was established and about how it works day to day. We'll look at why it's been successful and what lessons we've learnt along the way.

ABOUT THE AUTHOR

Briget Lander is a data and information management professional with many years' experience in commercial, not-for-profit and academic organisations. In her current role she encourages scientists to see that it's not only themselves who value the data they produce and supports them in taking good care of that data.

Harvesting Data: AgYields database a tool to support the New Zealand Agriculture Sector

Corresponding author: *carmen.teixeira@lincoln.ac.nz;
Presenting author: sonya.olykan@lincoln.ac.nz
derrick.moot@lincoln.ac.nz
Deepika.fernando@rezare.com

ABSTRACT

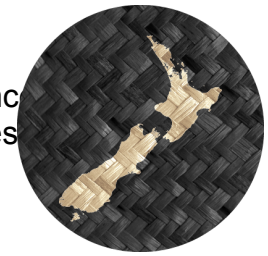
In the rapidly evolving landscape of technology, database platforms play a crucial role in collecting, managing, organizing and distributing data efficiently. Databases are structured collections of data that facilitate efficient storage, retrieval, and management of information. In the agriculture sector, data records, including crop and pasture measurements, are crucial for effective decision-making. They assist farmers to identify underperforming sites or paddocks due to factors like low-yielding species, poor fertility, drainage and nitrogen leaching or pests and diseases. By addressing these issues, farmers can improve productivity and reduce reliance on supplementary feed (Millar, 2023; Hoffman, 2024). Additionally, datasets can be combined with supplemented data to estimate pasture growth or contribute to pasture growth forecasting in software tools like Farmax (Farmax, 2024) and Overseer (Millar, 2023).

AgYields National Database (AgYields, 2024) is a relational model (Codd, 2007) created to provide a centralized and reliable source of agricultural yield data across different regions and farming systems in New Zealand. The database primarily focuses on pasture and crop yields (Moot et al. 2021) and is uniquely placed to accept and store academic published and unpublished data that consultants are requesting to enable them to advise farmers on crop and forage choices. It is freely accessible after a standard signup process. Relational model databases, such as AgYields, have become common due to their simplicity and robustness. The data is stored in tables (or relations), where each table consists of rows (records) and columns (attributes). Relationships between tables are established through foreign keys, enabling data normalization and reducing redundancy. This model supports powerful query languages, such as SQL, which facilitate complex data retrieval and manipulation. At a higher level, the architecture of a database can be summarized in three layers: the internal, conceptual, and external levels. The internal level describes the physical storage of data, including file organization and indexing methods that optimize performance. The conceptual level provides a unified view of the entire database, abstracting the complexities of data storage and presenting a logical structure that represents the data relationships. Finally, the external level defines how individual users interact with the data, often through customized views tailored to specific user needs, ensuring data security and accessibility (Alam et al 2013).

Understanding the structure of a database is crucial for effective data management. The interplay between different models and architectures informs the design choices that ultimately affect how data sre used by individual users or within an organization, highlighting the importance of aligning database structure with specific users and business needs and technological advancements.

This paper showcases the AgYields National database focussing on three main aspects: (i) the conceptual level presenting the database design and workflow from data input to the output and functionalities to attend the users' demand (ii) The organization of specific datasets within AgYields dashboard and how variables can be queried for relationships. (iii) the current and potential end users who are and will benefit from the Database.

ABOUT THE AUTHOR



Dr. Sonya Olykan is a Research Officer in the Department of Agricultural Scienc
Lincoln University, Canterbury, New Zealand. She is part of the Dryland Pastures

Group and contributes to the Hill Country Futures programme. Her research focuses on agronomy, soil science, and environmental science, with a particular interest in improving pasture management and sustainability. Dr. Olykan has a background in forest nutrition and sustainability, having previously worked for the Forest Research Institute (now Scion). She has published several key papers on topics such as subterranean clover management and sampling methods for diagnosing mineral deficiencies in pastures. She assisted on the conceptual phases of the AgYields database providing feedback on the Soil and Plant species components of the Database, the user interface experience and the navigation and online tutorials.

Research Officer – Department of Agricultural Sciences, Agriculture & Life Sciences Faculty, Lincoln University, Canterbury, New Zealand.
Links: https://drylandpastures.com/people/staff/dr-sonya-olykan/
https://www.youtube.com/playlist?list=PLr5_Hclr3uZVpxyfK9mIFZyXLty4Ml1fO
https://orcid.org/0000-0002-7482-8455

How can institutions enable researchers to use AI in conducting research appropriately and confidently?

l.armstrong@auckland.ac.nz, kyle.hemmimg@auckland.ac.nz, rina.hannaford@agresearch.co.nz, martinb@landcareresearch.co.nz

ABSTRACT / INTRODUCTION

Research institutions must embrace the opportunities offered by using AI in research or risk losing out on funding awards, failing to identify and explore innovative research, and not attracting research students and staff. Whilst individuals across institutions are doubtless using AI in research, these institutions need to develop and support cohesive responses to enable our diverse research community to i) more collectively and confidently weigh the benefits, limitations and risks of AI tools and technologies, ii) select and use appropriate AI tools to conduct research, and iii) engage with research students and colleagues in understanding the broader and evolving landscape of AI across the research sector.

In this session, we will facilitate an open discussion to draw input from the audience on the following questions:

- *In the context of responding to, adopting, or driving use of AI in conducting research, what are institutions experiencing as the researcher development needs?*
- *Are there priority areas?*
- *What are institutions doing about those needs?*
- *What barriers/enablers do institutions have in meeting those needs?*
- *What, if anything, might it be useful to share or collaborate over?*

This session is aimed at those in the research sector interested in enabling appropriate and confident use of AI in conducting research, including researchers, research support staff, eresearch service providers, institutional administrators, research librarians, cybersecurity staff, etc.

ABOUT THE AUTHOR(S)

Laura Armstrong
Laura is the eResearch Engagement Lead for Waipapa Taumata Rau, University of Auckland. She collaborates to engage with the research community to raise awareness and use of modern technologies and tools to advance research. Her focus in recent years has been maturing research data management through researcher development, service delivery and policy.
http://orcid.org/0000-0003-2370-3924

Dr Kyle Hemming
Kyle is a Senior eResearch Engagement Specialist in the Centre for eResearch, Waipapa Taumata Rau, University of Auckland. With a PhD in Conservation from the University of Canberra, he is experienced in research coding, aiming to address pressing global issues through innovative solutions. Kyle is passionate about upskilling researchers in new tools and technologies to enhance the impact of their work. His interests include data analysis, AI, and fostering collaborative research environments.

Dr Rina Hannaford
Rina is a data scientist at AgResearch who is involved in the organisation's emerging digit
development programme. Rina is a certified Carpentries instructor and has a backgroun

Dr Brent Martin

Brent is a senior data scientist and machine learning specialist at Manaaki Whenua – Landcare Research. His 38-year career spans both academic research as a senior lecturer at Canterbury University, as well as software engineering and R&D roles in various commercial companies, from local software house Jade to Google NY. For the past 12 years Brent has focused on tackling pressing environmental issues, including 7 years helping companies and individuals reduce their GHG emissions as the software architect at Toitū Envirocare. Brent's research in AI and machine learning includes developing new ML classification algorithms, applying ML to real-world problems such as electricity demand forecasting and internet search engines, research and development in Intelligent Tutoring Systems, developing social network analysis and anomaly detection techniques for criminal investigation, and applying deep learning to environmental computer vision problems. At Manaaki Whenua, Brent has applied deep learning to problems for a range of clients including government departments such as MfE and MBIE, and regional councils throughout New Zealand, as well as internal projects aimed at automating and upscaling some of the services Manaaki Whenua already provides. More recently this research has included applied generative AI.

HPC for Quantum Simulation: Challenges and Learnings

Peter McGonigal
XENON Systems, Melbourne, Australia
petermc@xenon.com.au

ABSTRACT

The Pawsey Supercomputing Research Centre in Perth is home to one of the most powerful research HPC installations in the Southern Hemisphere.
Pawsey is also home to the first room-temperature quantum computer through a partnership with Quantum Brilliance. To support this work, Pawsey build a small HPC cluster using latest in GPU and high-speed storage to simulate quantum computations to enable programming and development on traditional silicon.
This talk will outline the technical solution, performance metrics achieved, and key learnings demonstrating how a very fast HPC solution can be built with a small number of nodes to support research into quantum computing.

ABOUT THE AUTHOR

Peter McGonigal joined XENON Systems in 2016 as a Solution Architect. Prior to XENON Systems, Peter was a Senior Systems Engineer at SGI (Silicon Graphics) with a background in HPC and complex Data Management systems.

Peter is a motivated and experienced professional with a deep understanding of High Performance Computing (HPC), high-speed big capacity data storage technologies, and research and technical computing workflows and environments. Proven ability to design, implement, and manage complex HPC and data storage systems, ensuring smooth operation and optimal performance for diverse research and technical computing needs. Proficient in collaborating with researchers and technical users to comprehend their requirements and formulate efficient solutions.

Peter has worked with the largest HPC customers and data sets in Australia, across Pawsey and CSIRO. https://www.linkedin.com/in/pmcgonigal/

Hybrid HPC: can we finally manage our data easily and consistently between on-prem and cloud?

Peter McGonigal and Adrian Torrie
XENON Systems, Melbourne, Australia
petermc@xenon.com.au and adriant@xenon.com.au

ABSTRACT

Public cloud use in a hybrid HPC environment remains a challenge: compute is easy, but data access comes with complications.
Storage vendors have implemented various approaches to the problem of bursting data to the cloud. Some use global namespaces to stretch the filesystem others use snapshots and others shared repositories.
In this talk we explore these different approaches: their strengths, weaknesses and scenarios where they best apply.

ABOUT THE AUTHORS

Peter McGonigal joined XENON Systems in 2016 as a Solution Architect. Prior to XENON Systems, Peter was a Senior Systems Engineer at SGI (Silicon Graphics) with a background in HPC and complex Data Management systems.

Peter is a motivated and experienced professional with a deep understanding of High Performance Computing (HPC), high-speed big capacity data storage technologies, and research and technical computing workflows and environments. Proven ability to design, implement, and manage complex HPC and data storage systems, ensuring smooth operation and optimal performance for diverse research and technical computing needs. Proficient in collaborating with researchers and technical users to comprehend their requirements and formulate efficient solutions.

Peter has worked with the largest HPC customers and data sets in Australia, across Pawsey and CSIRO. https://www.linkedin.com/in/pmcgonigal/

Adrian Torrie, of Ngāti Porou descent, is a skilled solutions architect with broad experience building and maintaining data systems from the ground up.

He has broad skills across all IT domains, with specific depth in automating data analytics/machine learning/artificial intelligence solutions in hybrid environments, along with the necessary skills for training others in this specialty.

Adrian's last role had a primary focus of enabling self-service, via Platform Engineering, for Data Scientists and Data Engineers, allowing rapid iteration of models and their release.

He excels in delivering secure solutions for large scale, high performance computational requirements using modern automation approaches.

https://www.linkedin.com/in/adriantorrie/

Innovation through diverse approaches, how we adapt and evolve

Jun Huh, NeSI, jun.huh@nesi.org.nz
Claire Rye, NeSI, claire.rye@nesi.org.nz
Thomas Berger, NeSI, thomas.berger@nesi.org.nz
Nathalie Giraudon, NeSI, nathalie.giraudon@nesi.org.nz
Georgina Rae, NeSI, georgina.rae@nesi.org.nz

ABSTRACT

NeSI has evolved its ways of working and transformed through rich partnerships for collaboratively developing solutions for various research communities. NeSI has embraced diversity in its team culture to address equally diverse challenges facing the eResearch communities. This talk showcases some of the key projects as case studies, highlighting different ways of working and how NeSI has applied product management and product engineering practices. These include both embedding them into the team and sharing the practices with the partners to enable a collaborative journey that is user focussed, iterative and transparent.

With Rebase, NeSI's new infrastructure provisioning, we have had the opportunity to apply and improve all of these practices, as the new infrastructure brings complexity to power the ever-growing diversity of the research communities. Much of the infrastructure provisioning journey has also been learnt from NeSI's ongoing partnership with AgResearch, where we have collaborated to set up the eResearch Infrastructure for the researchers in AgResearch.

As far back as 2018, NeSI embarked on development of new services, a step away from its core HPC service. These include the National Data Transfer Platform and JupyterHub. NeSI has then embarked on a partnership with Genomics Aotearoa for developing the Aotearoa Genomic Data Repository. Here, the journey started off as discovery-oriented putting NeSI onto a path of learning Kaitiakitanga, Māori Data sovereignty, and data standards for genomics, alongside cross institutional agile processes. The partnership continues to this day, where the shape of the collaboration has transitioned from pioneering into more increased maturity.

ABOUT THE AUTHORS

Jun Huh is a Product Manager at NeSI. Jun brings his experience from start-up industries into the field of eResearch. He has been involved in genomic data management related projects for the past 2-3 years, to help build a data repository system in partnership with Genomics Aotearoa, and more recently, prototyping for Rakeiora Pathfinder project, which focuses on enabling research while retaining full visibility and control of data to the indigenous communities.

Claire Rye is a Product Manager at New Zealand eScience Infrastructure (NeSI) based out of the University of Auckland. She is responsible for the National Data Transfer Service and works across the Aotearoa Genomics Data Repository and Rakeiora Pathfinder projects and looking at research data management and data lifecycle more generally across NeSI, and is a co-chair for the World Data System Early Career Researchers network. Claire holds a PhD in organic chemistry and has spent the last 11 years working in the UK in a variety of research settings.

Thomas Berger is a Product Manager at NeSI. Thomas has been focused on improving the user experience of NeSI services by continuously improving my.nesi, a researcher-facing portal, and also the NeSI JupyterHub. Thomas brings to NeSI over a decade of experience in product management skills with strong user focus and value driven approach.

Nathalie Giraudon holds an electrical engineering degree from ENSEEIHT, École nationale supérieure d'électrotechnique, d'électronique, d'informatique, d'hydraulique et des télécommunications (French engineering school based in Toulouse). After an exchange year to McGill University, Montreal, Canada, Nathalie has worked more than 30 years in telecommunications industry (Nortel Networks, Tait Communications). She is passionate about system design and business analysis. With her knowledge in product research and development, she is now the product engineering lead of the Collaboration and Integration team at NeSI.

Georgina Rae is the Science Engagement Manager at NeSI where she ensures that NeSI is supporting NZ's researchers and research priorities through meaningful partnerships and user-driven approaches. Prior to NeSI she has worked in molecular biology and intellectual property.

Integrating information flow control within programming languages and operating systems to effect data sovereignty tracking

David Eyers
University of Otago
david.eyers@otago.ac.nz

ABSTRACT

Information flow control (IFC) is a computer security approach that involves attaching labels to sensitive data and tracking these labels as they pass through the transformations applied by software systems. Within programming languages and operating systems, IFC can be used to enforce mandatory access control (MAC), where participants (e.g., users and applications) within the software system also carry labels, and various security rules apply by default, based on comparisons between the IFC labels of participants and data items. This type of security is commonly applied within military and government intelligence sectors, with rules such as “no write down”, that indicates information cannot be declassified without an explicit permission to do so, and “no read up” that indicates that participants without permission cannot read data that is labelled as being more sensitive than their own level. However most computer security today employs an alternative to MAC, namely discretionary access control (DAC), where the creator of a data item can arbitrary change its permissions. In this work we present early-stage experimentation with IFC technologies for carrying labels that support data sovereignty constraints, e.g., for the support of Maori data sovereignty. While most mainstream software systems, such as relational databases, have little or no support for IFC protections, there is a growing need to systematically handle sensitive data throughout its lifecycle, and IFC provides a promising means for achieving secure data handling within OS and programming language contexts. Here we focus more on the mechanisms by which labels are tracked and rules are specified to use the labels for protection, with a view to how these low-level protection mechanisms might support interaction with emerging technology such as data spaces.

ABOUT THE AUTHOR

David Eyers is a professor in the School of Computing at the University of Otago. He has broad research interests covering security, AI, cloud, operating systems, networks and distributed systems. One of his research themes involves investigating how emerging, low-level computer security technologies can support high-level security goals that are meaningful to the users of these computer systems.

Introducing UC’s new Research Platform

François Bissey, eResearch Consultant, University of Canterbury francois.bissey@canterbury.ac.nz
Jamie Hart, Product Manager, University of Canterbury Jamie.hart@canterbury.ac.nz

ABSTRACT

The future of research computing at UC is launching in February 2025. With a mandate to enable relevant, distinctive and world class learning, research, and collaborative experiences to a global audience, eResearch at UC is a crucial enabler for UC’s Strategic Vision 2030. The UC Research Compute Cluster reached end of life in 2024, and as part of the UCs Digital Transformation’s Cloud Programme, a project to invest in a new private cloud was launched. Following a RfP process, UC has partnered with Sempre, and HPC Now! to deliver the highly anticipated research platform for UC researchers and collaborators.
As at the date of the eResearch NZ 2025 conference, UC will be on the cusp of delivery. We will give a sneak peek of the platform and share what is on the horizon.

ABOUT THE AUTHORS

Dr François Bissey is a particle physicist that knows too much about computers. After two postdoctoral positions, he took a position as supercomputing consultant at the university of Canterbury as part of the original pool of NeSI staff.
He is currently serving as a eResearch consultant as part of Digital Services. His mission being to help researchers with their digital problems and to champion their interest to the wider Digital Services unit.

Jamie Hart is a Product Manager in the Digital Services Unit, leading the Research Experience team. Jamie’s team supports UC's digital research infrastructure and provides eResearch services to UC researchers across the research lifecycle. With a background in Law, academic publishing and the performing arts, Jamie thrives on building strong relationships with stakeholders across UC and beyond and understanding the needs of the researcher in order to drive positive change. A key focus for the Research Experience team at UC is on building the eResearch catalogue of services at UC.

Investigating the advantages and disadvantages of available open-source frameworks for prototyping data-based applications

Rebekah Au, Phil Davies | University of Canterbury
Lei Chen, Annabelle Bos | Environmental Science and Research
rebekah.au@esr.cri.nz

ABSTRACT

The Data Science team at ESR focus on exploring and developing innovative solutions that can then be passed on to clients or other areas of the organisation to use and maintain. Data Engineers employed within this team create platforms which enable users to easily interact with and interpret data and results. Traditionally these have been developed in Tableau or R Shiny, but more recent projects have been built in Streamlit or React.

The shift from R Shiny to Streamlit was initially made because Streamlit was supported by Snowflake, which the team were moving towards as a data management platform, and applications could be created and deployed rapidly. ESR are now using a High-Performance Computing cluster for data management and model development, so other options beyond Streamlit could be considered. While Streamlit is easy to use, it does have several disadvantages that may mean that another framework could be a better choice for certain projects.

The goal of this research project was to compare Streamlit to potential alternatives. Options such as Plotly Dash and Python Shiny were compared on factors such as the overall look and feel of the application, complexity of the visualisations, optimisation of performance, the use and persistence of input widgets, and the ability to incorporate data modelling. A decision-support tool was developed to aid the Data Engineering team in determining which package to use for their prototype applications. As part of the exploration process, a template application was built in several frameworks for comparison purposes. Future research could use this template to explore further options.

ABOUT THE AUTHOR

Rebekah Au is a Data Science intern at Environmental Science and Research who is currently working on an industry-based student project with the Data Science team. This project is a contribution towards a Master of Applied Data Science from the University of Canterbury. During her internship, Rebekah has developed dashboards and formed strategies to visualise and model time series and geospatial data.

Learning from Industry Data Science Platforms: A Leapfrog Opportunity for eResearch in Aotearoa

Adrian Torrie
Xenon Systems, Melbourne, Australia
adriant@xenon.com.au

ABSTRACT

Learn from, and leverage, industry's investment of time and capital to revolutionise eResearch capabilities.

Aotearoa's eResearch community has a unique opportunity to leapfrog legacy approaches by adopting proven industry practices and tools in modern platforms. This session explores how implementing modern cloud-native approaches can transform research capabilities whilst embedding best practices for data and model governance and drive researcher engagement with self-service tools.

Drawing from industry experience in data engineering, data science, and platform engineering, across hybrid cloud-native environments, we outline practical steps for modernisation through four key areas:

- **Platform Modernisation**: Combining Kubernetes with HPC-specific schedulers like YuniKorn and high-performance storage solutions enables research teams to maintain HPC performance while gaining cloud-native flexibility.
- **Automation**: Implementing standardised patterns to data engineering, model deployment, monitoring, and lifecycle management ensures reproducibility and reliability. Modern tools enable streamlined version control and automated deployment processes in a cloud-native fashion.
- **Workforce Development**: Building cloud-native critical skills enables researchers to leverage industry best practices in platform engineering, data engineering, data science, and includes them into the wider cloud-native ecosystem and worldwide community. This enhances both individual career mobility between research and industry while strengthening research institutions' capability for engagements with industry.
- **Stewardship**: Modern platforms enable granular control over location, access, and governance, ensuring sovereignty of choice and practices.

Practical examples will demonstrate:

- Converting traditional workloads and workflows to cloud-native approaches
- Implementing platforms with sovereignty considerations
- Establishing automated workflows leveraging proven industry patterns
- Optimising performance in hybrid environments
- Modern tools for researchers interested in utilising AI and machine learning

By adopting industry-hardened practices, Aotearoa's research sector can build future-ready platforms and people that support innovation and ensure sovereignty of world-class research outcomes. This transformation creates an inclusive, collaborative research environment aligned with best practices, building resilience for our digital future.

Thanks to industry leaders, mentors, and the Aotearoa eResearch community for their commitment to advancement and collaboration. Special acknowledgement to those working to ensure research infrastructure supports inclusive and equitable access, and sovereign practices.

Key references will include reference architectures, best practices, design patterns, and frameworks.

ABOUT THE AUTHOR

Adrian Torrie, of Ngāti Porou descent, is a skilled solutions architect with broad experience building and maintaining data systems from the ground up.
He has broad skills across all IT domains, with specific depth in automating data analytics/machine learning/artificial intelligence solutions in hybrid environments, along with the necessary skills for training others in this specialty.
Adrian's last role had a primary focus of enabling self-service, via Platform Engineering, for Data Scientists and Data Engineers, allowing rapid iteration of models and their release.
He excels in delivering secure solutions for large scale, high performance computational requirements using modern automation approaches.

https://www.linkedin.com/in/adriantorrie/

Leveraging Computer Vision and Hybrid Transfer Learning for Deceased Bird Species Classification

Zhuo TANG, James Williams | University of Canterbury
Richard Dean, Jiawei Zhao | ESR
zta38@uclive.ac.nz

ABSTRACT

Recent advancements in deep learning, particularly in computer vision, have driven significant progress in image classification tasks using CNN-based and transformer architectures. These models are widely applied in computer vision due to their robustness and high accuracy. In response to recent global outbreaks of Highly Pathogenic Avian Influenza (HPAI) and the associated risks to New Zealand's native birds, poultry industry, and public health, our team aims to explore the deployment of computer vision technologies to enhance New Zealand's avian flu pandemic preparedness.
While computer vision models for classifying live bird species have achieved notable success, often exceeding 90% accuracy through fine-tuning of CNN and transformer-based architectures, they face challenges when applied to images of deceased birds. These images tend to be noisier and lack the distinct features found in live bird images. Our solution aims to develop a fast, lightweight model with minimal GPU and memory requirements, capable of classifying both dead and live bird species and deployable on mobile devices.
To address this challenge, we collected and labelled 1,200 images of dead birds from New Zealand, using EfficientNetB2 and Vision Transformer (ViT) architectures as the backbone of our model.
Our presentation will cover the following objectives:

1. Evaluate the performance of EfficientNetB2 and ViT architectures when trained directly on images of deceased birds.
2. Explore a hybrid approach that combines CNN and transformer-based transfer learning with traditional classifiers for deceased bird species classification.

We will further elaborate on why this hybrid approach is particularly suited for accurately classifying both deceased and live bird species. Our methodology and results will demonstrate how this hybrid transfer learning approach is ideal for mobile applications. Additionally, we will showcase a demo of our app, highlighting its rapid, accurate, and lightweight mobile-based model deployment.

ABOUT THE AUTHOR

Zhuo TANG is a master's student in Applied Data Science at the University of Canterbury, with research interests cantered on Data Science, the Application of AI in GI science, and the intersections between these fields.
Richard Dean is a senior data scientist in ESR's core data science team. He has over 20 years' experience working with health, forensic and environmental data sets, specialising in the development of novel data tools and visualisation techniques. He currently leads a research programme looking at the use of computer vision for rapid diagnostics and is responsible for the development of ESR's digital twin user interface.
Jiawei Zhao is ESR's machine learning engineer with a doctoral degree in computer science, specializing in natural language processing, computer vision and speech recognition. His research interests centre around advancing the capabilities of these domains, pushing the boundaries of artificial intelligence.

Keynote: Local Contexts: Supporting Iwi Rights and Interests in Data and Research

Dr Janette Hamilton-Pearce (Te Whānau a Apanui)
Chief of Staff, Local Contexts

ABSTRACT

This presentation will focus on introducing the Local Contexts tools for supporting Iwi rights and interests in data and research practices. Institutional research practices in Aotearoa are not neutral. These research practices take on the colonial practices to promote Māori as the 'Dying Race' driven by Issac Featherston since 1856. To counter that narrative, you will learn about the Local Contexts Traditional Knowledge (TK) and Biocultural (BC) Labels and Notices initiative, and how they can be implemented within research and publication processes.

ABOUT THE AUTHOR

Dr Janette Hamilton-Pearce is a descendent of the Te Whānau a Apanui iwi of Aotearoa New Zealand. Dr Hamilton-Pearce specialises in supporting Indigenous communities in managing intellectual and cultural property rights and data within information systems by ‘Tagging the World with Traditional Knowledge (TK) and Biocultural (BC) Labels’ at Local Contexts.

She is the Chief of Staff at Local Contexts with over 28 years experience as an academic in information systems for Indigenous communities. She obtained her Doctor of Philosophy in 2009 from the School of Computing and Mathematical Sciences, Auckland University of Technology (AUT).

Keynote: Maximising AI Performance with Minimal Data

Prof Dr Amanda S Barnard AM
School of Computing, Australian National University

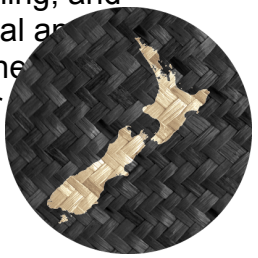
ABSTRACT

Machine learning is a powerful way of discovering new science, designing new systems, and screening new possibilities. However, moving from a promising prediction to a practical strategy often requires more than just an instructive correlation. Understanding how machine learning uses data to make decisions can often be critical. Explainable artificial intelligence (XAI) is an emerging field in computer science based in statistics that can augment scientific machine learning workflows. XAI can be used as a forensic analysis technique to understand the consequences of data, model, and application decisions, or as a model refinement method capable of distinguishing important information. This approach is often applied to the feature space to explain the how the independent variables contribute to the predictions, using tools such as feature rankings to identify useful or nuisance variables. However, an alternative approach is to apply similar methods to the instance space, and identify influential or unproductive data points. In this presentation we investigate these opportunities, by exploring high-dimensional patterns structured (tabular) materials data sets in behavioral space instead of the feature space. Behavioral vectors are used to represent the contribution of an instance to an interpretable quantity (such as a material property), which complement more conventional interpretations of machine learning models. We make use of Shapley values to decompose well-studied summary statistics of the data to give rise to different interpretable clusterings and modalities compared to the original data. Our approach is efficiently demonstrated qualitatively and quantitatively over three material data sets and uncovers hidden characteristics that may aid the data analysis process. Once sets of influential instances (materials) have been identified, we also decompose the residuals of regression with respect to the data instances, to determine the effects of each individual instance on the model and each other. This provides a model-agnostic method of identifying instances of interest, which can determine the appropriateness of the model and data in the wider context of a given study, as well as providing unique material insights.

ABOUT THE AUTHOR

Professor **Amanda Barnard** is one of Australia's most acclaimed computational scientists, renowned for her expertise at the intersection of computational modeling, high-performance computing, and applied machine learning and artificial intelligence (AI). She holds a BSc (Hons) in Applied Physics (2000), a PhD in Theoretical Condensed Matter Physics (2003), and a DSc (2020) from RMIT University.

Following her studies, Amanda held distinguished international fellowships, including a Postdoctoral Fellowship at the Center for Nanoscale Materials at Argonne National Laboratory (USA, 2003–2005) and as the Violette & Samuel Glasstone Fellow at the University of Oxford (UK), alongside an Extraordinary Research Fellowship at The Queen's College (2005–2008). Upon returning to Australia, she served as an ARC QEII Fellow, Office of the Chief Executive (OCE) Science Leader, and later Chief Research Scientist in Data61 at CSIRO (2009–2020). In 2020, Amanda joined the Australian National University, where she is now a Senior Professor and Deputy Director of the School of Computing. With over two decades of experience in high-performance computing, computational modeling, and informatics, she serves on multiple institutional boards and has earned 11 national and international awards spanning five scientific disciplines. She is also a Fellow of the Institute of Physics, the Royal Society of Chemistry and the Australian Computer



recognition of her contributions to science and education, Amanda was made a Member of the Order of Australia (AM) in 2022.

Modernising Scientific Software and Workflows for Earthquake Hazard Research

Jake Faulkner[1] (presenting author), Sung Bae[2], Brendon Bradley[3], Robin Lee[4], Cesar Pajaro[5], Joel Ridden[6], Andrew Ridden-Harper[7], Claudio Schill[8].
University of Canterbury
[1] jake.faulkner@canterbury.ac.nz, [2] sung.bae@canterbury.ac.nz, [3] brendon.bradley@canterbury.ac.nz, [4] robin.lee@canterbury.ac.nz, [5] cesar.pajaro@canterbury.ac.nz, [6] joel.ridden@canterbury.ac.nz, [7] andrew.ridden-harper@canterbury.ac.nz, [8] claudio.schill@canterbury.ac.nz

ABSTRACT

This presentation documents a transformative modernisation of the scientific workflow currently used for earthquake-induced ground motion simulations and hazard assessment in New Zealand. Starting with a rigid legacy codebase, we have adopted the latest standards of software engineering and computational science to develop a unified workflow. The new workflow architecture is flexible enough to accommodate custom workflows for individual researchers, and the next iteration of the 'Cybershake NZ' project, simulating thousands of rupture scenarios from the NZ National Seismic Hazard Model (NSHM). Through our modernisation, we have enhanced scientific reproducibility by introducing a containerised software stack, adopting Cylc for workflow management, and leveraging software engineering practices such as code review and GitHub Actions for rapid iteration through continuous testing.

Our new workflow also addresses new scientific challenges in ground motion simulations from the latest NSHM – specifically, multi-fault rupture simulations involving many faults spanning the length of the country. We have also enabled the development of near real-time simulations of recorded earthquakes integrating with GeoNet directly, opening new possibilities for rapid earthquake analysis to inform disaster response, as well as significantly enhancing public science communication. This talk will share the lessons learned in enabling these advances, and outline the future changes aimed at catalysing research through software.

We also acknowledge NeSI for the computational resources enabling our work.

ABOUT THE AUTHOR

Jake Faulkner is an early-career research software engineer at the University of Canterbury. His work involves developing scientific code and providing support for research within QuakeCoRE: The NZ Centre for Earthquake Resilience. With an undergraduate degree in Computer Science and a Master's in Mathematics, Jake specialises in applying computational methods to scientific research. Jake is currently leading the development of the new scientific workflow for ground motion simulations.

My Preciousss – Making sensitive metadata resilient

Chris Seal, James Love, Yvette Wharton and Tom Laurenson
The Centre for eResearch, University of Auckland
c.seal@auckland.ac.nz, james.love@auckland.ac.nz, y.wharton@auckland.ac.nz,
tom.laurenson@auckland.ac.nz

ABSTRACT

While data is typically held in a resilient storage system (often with 11+ 9s of reliability), metadata is more often held in more ephemeral spaces, such as databases. While these systems are designed to be resilient, if the metadata is lost, the data becomes difficult, if not impossible, to identify and use. A better solution is to store the metadata alongside the data, and there is an open standard, research object crate (RO-Crate), that fulfils this well. While RO-Crate meets the majority of the metadata needs, it is a plain text format that does not sufficiently protect metadata fields that are sensitive, such as identifiable data, for example. This use case has come to light as part of the instrument data service strategic project that the Centre for eResearch at the University of Auckland has led. During the project, an approach using the Pretty Good Privacy (PGP) standard has been developed. This approach allows sensitive metadata fields to be encrypted against individuals and/or service account public keys and stored as encrypted strings within the RO-Crate. When the metadata is read into memory, if the user provides an appropriate private key, the encrypted section is decrypted and read back into the metadata structure. Decrypted metadata is only retained in memory, and is accordingly ephemeral. When the RO-Crate is written to disc, the scripts developed re-encrypt the data and write out the encrypted copy.

Nectar in New Zealand: Seven Years On

Sean Matheny, Martin Feller,
Waipapa Taumata Rau, University of Auckland
s.matheny@auckland.ac.nz, m.feller@auckland.ac.nz,

ABSTRACT

In February 2017, at eResearch New Zealand, ministers from New Zealand and Australia committed to enhanced collaboration across the Tasman in the eResearch domain, with one key goal focusing on shared projects and technologies.[1] This significant step forward enabled the Centre for eResearch at the University of Auckland to activate the Auckland node of the Nectar Research Cloud the following year, with substantial support from the ARDC.

Over the past seven years, the landscape of cloud research computing has transformed dramatically. We have seen the emergence of numerous new tools, advancements in AI, the development of new workflows, and the need for new skills. There has also been an increased emphasis on cybersecurity, data sovereignty, and environmental sustainability, alongside a substantial boost in overall capability.

This presentation will highlight how we have adapted to these changes, share key lessons learned, and discuss our future directions.

ABOUT THE AUTHORS

Sean Matheny began his career working in the physics department of the largest radiation oncology treatment provider in the U.S. There, he helped to improve QA, dosimetry, and treatment planning and delivery via technology, machine learning, and automation. For the last 8 years, he's primarily worked in roles that have enabled research cloud platforms and HPC (high performance computing) in both the Centre for eResearch (University of Auckland), and NeSI. Currently, he helps operate the Auckland node of the Nectar Research Cloud, at the Centre for eResearch.

Martin Feller Since moving to New Zealand Martin Feller has worked at the Centre for eResearch at the University of Auckland in various roles. He is currently leading the Platform and Services Team, which is looking after the operations of the Nectar OpenStack platform, other backend services, automation and reporting.

1

Opportunities in Horizon Europe for NZ eResearch

John Cater
University of Canterbury
cluster4ncp@mbie.govt.nz

ABSTRACT

Horizon Europe is the largest public research funding programme in the world and aims to address global challenges including climate change, digital transformation, and sustainable development. New Zealand has recently associated with the programme for Pillar 2, which supports multi-national consortia of researchers and businesses (including SMEs) to develop science and technological solutions.
This seminar will focus on the specific opportunities available to New Zealand researchers, companies, and innovators within Horizon Europe in the Digital, AI and HPC areas. Additionally, the talk will provide practical guidance on navigating Horizon Europe's application process, eligibility criteria, and key calls for proposals in the 2025-2027 Work Programme.

ABOUT THE AUTHOR

Professor John Cater is the New Zealand National Contact Point (NCP) for Horizon Europe, Cluster 4 (Digital Industry & Space).

ORCID in New Zealand: Enhancing Researcher Discoverability, Collaboration, and Impact

Dr Tifany Oulavallickal
Kaiārahi ORCID | ORCID Lead
tifany.oulavallickal@royalsociety.org.nz

ABSTRACT

This lightning talk introduces the role and integration of the Open Researcher and Contributor ID (ORCID) within the research community in Aotearoa New Zealand. I will highlight how ORCID enables researchers to manage their scholarly contributions through a persistent digital identifier, and provide an overview of the NZ ORCID Hub, which facilitates Consortium members to productively engage with ORCID regardless of technical resources. All enquiries related to ORCID, the ORCID Hub, and their implementation are welcome, both during and after the conference.

ABOUT THE AUTHOR

Tifany Oulavallickal is the ORCID Lead at the Royal Society Te Apārangi, supporting the implementation of ORCID for researchers and organizations. With a multidisciplinary background, including a Master's in Engineering and a Ph.D. in Biochemistry, Tifany brings a comprehensive understanding of New Zealand's research landscape to the role.

Out of the squares and into the circles: Visualising Ecological Interactions with Circos

Senzo Miya, Mateus Detoni, Manpreet Dhami
Manaaki Whenua Landcare Research
miyas@landcareresearch.co.nz

ABSTRACT

Circos is an open-source software package designed for visualising data and information (1) that traditional graphs can't easily handle. It visualises data in a circular layout, which makes it ideal for exploring interactions between objects or positions. It was originally developed to be used by bioinformatics scientists for visualisation of genomic features and sequencies. However, it has since found applications in various fields and studies, including analysis of migration flows (2) and analysis of transcribed political speeches (3).
In our research, we used Circos to investigate interactions between pollinators, microorganisms and plant species. Although the results obtained were inconclusive, they highlighted areas needing further analysis. Circos has proven to be a great tool for identifying promising areas for deeper data exploration.

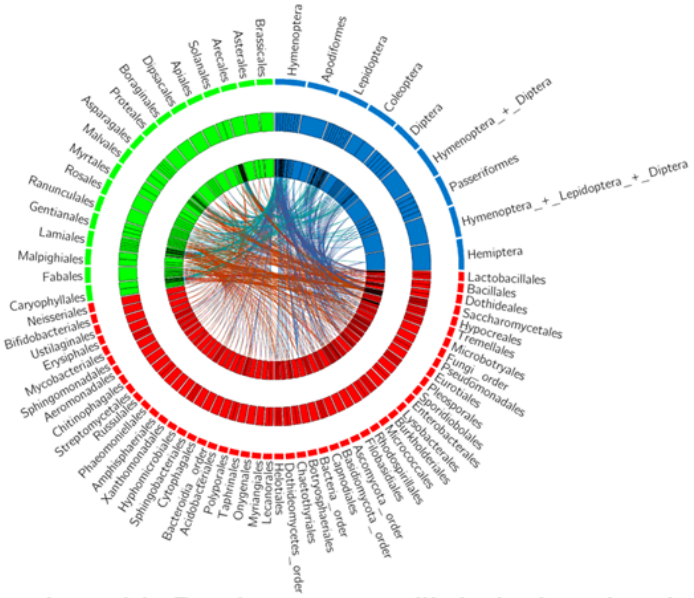


Figure 1: Circos plot with Bezier curves (links) showing interactions between pollinator, microorganism and plant species.

ABOUT THE AUTHOR

Senzo Miya is originally from South Africa and studied material science before moving to New Zealand for a post-doctoral research position at the University of Canterbury. After completing his post-doctoral program, he worked at Statistics New Zealand, where he gained insight into the critical role of data in decision-making.
Following his time at Statistics New Zealand, Senzo pursued a postgraduate diploma in applied data science. After graduating, he secured a position as a research data engineer at Manaaki Whenua Landcare Research, where he has been working for approximately two and a half years.

Petabyte Partnerships: Managing Data for a Collaborative Science Future

Ben Taylor
AgResearch
Ben.Taylor@agresearch.co.nz

ABSTRACT

Managing data at Petabyte scale is hard. At AgResearch, with over 300 million files and 2PB+ this presents unique challenges in terms of supportability, cost, disaster recovery, user experience to name a few. The problem is about to get exponentially harder with a proposed CRI merger where there will be a need to amalgamate multiple large storage systems.
This BoF seeks to explore ways forward, both from an infrastructure perspective and change management. Throwing money at the problem by leveraging bigger and bigger filesystems seems untenable, and a paradigm shift may be in order. Numerous technologies exist that can assist with making everyone's life easier (such as Object Storage), but these all have their own challenges, and lets face it, who likes to archive data?
Come join me for a collective sharing of war stories, then lets put our heads together and imagine a better future!

ABOUT THE AUTHOR

Ben Taylor is the eResearch, Infrastructure and Security Manager at AgResearch with over 20 years of experience. By day (and sometimes by night), he juggles a range of responsibilities to ensure AgResearch's researchers have the infrastructure and services they need to feed their ever-growing data appetite. For his sins, he also tries to balance this with the need to protect information and data. Easy right?

Population Simulation to Optimise Study Designs and Estimate Polygenic Disease Risk/Resilience in Aotearoa Māori Populations

Dr Alastair Lamont, Professor Mik Black, Associate Professor Philip Wilcox
University of Otago
alastair.lamont@otago.ac.nz, mik.black@otago.ac.nz, philip.wilcox@otago.ac.nz

ABSTRACT

For commonly occurring polygenically inherited conditions such as gout, type 2 diabetes, and cardiovascular conditions, disease risk/resilience estimates have most often been derived from GWAS (genome-wide association studies). Such studies require large sample sizes ($n > 10^4$ participants) genotyped with 10^4-10^7 DNA markers.
These datasets often do not include indigenous peoples, who can have important genetic differences from more commonly represented populations of predominantly European descent. Moreover, existing datasets from Māori (and Pasifika) domiciled in New Zealand are few, and those that could be utilised consist of fewer than two thousand individuals - and thus are underpowered for clinically accurate disease risk/resilience prediction. In addition, establishing sufficiently large GWAS is unlikely in Aotearoa/NZ because of substantive costs associated with generating genotypic data and reluctance of many Māori to participate in such studies.

In order to offset further health inequities arising from a lack of Māori-specific DR prediction models, new studies are required. Such studies require both (a) optimal designs that incorporate known genetic relationships on non-genotyped as well as genotyped individuals, and (b) analytical methods that more accurately predict phenotype than GWAS-based methods such as polygenic risk scores.

We have used a population simulator (SLiM) to model genetic structures of Māori communities (i.e., whānau/hapū/iwi), incorporating estimates of effective population sizes prior to European admixture, as well as post-colonisation admixture with Europeans. Through NeSI, we are using these simulations to explore what features of study design and analytical methods lead to optimal disease risk/resilience prediction. I will illustrate and present current results on this.

ABOUT THE AUTHORS

Dr Alastair Lamont is a Postdoctoral Fellow with the Department of Mathematics and Statistics at the University of Otago. He is interested in applying statistical methodologies to improve prediction of human genetic traits, particularly in indigenous peoples.

Professor Mik Black is a statistician, bioinformatician and data scientist whose research focuses on the development of methods for the analysis of genomic data, with a strong emphasis on cancer and other human diseases. A common theme is the use of techniques that allow high-dimensional and often very disparate data sets to be combined in ways that provide new insights into disease development and progression. In addition to his own work,

Mik has been heavily involved in establishing national research infrastructure in high performance computing through the NZ eScience Infrastructure, and in genomics and bioinformatics through Genomics Aotearoa, where he is Co-Director and chair of the Bioinformatics Leadership Team.

Associate Professor Phillip Wilcox's Māori tribal affiliations are Ngāti Rakaipaaka, Ngāti Kahungunu ki te Wairoa, Rongomaiwahine and Te Aitanga a Mahaki. He is based in the University of Otago's Department of Mathematics and Statistics, and has experience in applied genomics and statistical genetics, as well as engagement with indigenous communities regarding gene technologies. He is also an Affiliate of the University of Otago's Bioethics Centre, and is the current convenor of MapNet, a NZ-wide collective of gene mapping scientists, and led the Virtual Institute for Statistical Genetics from 2008 to 2013. He is also a Deputy Director of the Maurice Wilkins Centre. For over 20 years he has worked in the interface of genetic sciences and Te Ao Māori, and co-leads four genomics-based projects focussing on Māori health. A/Prof Wilcox has also worked on genetics of plant species (particularly forest trees) and Māori health. He also co-initiated the Summer Internship of iNdigenous peoples in Genomics Aotearoa (SING-Aotearoa), and was a member of the Health Research Council of New Zealand's Ethics Committee which oversees New Zealand's institutional and regional ethics committees.

Rebasing infrastructure and the friends you make along the way

Nick Jones, nick.jones@nesi.org.nz

ABSTRACT

NeSI has rebuilt its infrastructure platform over the last year, from a signed RFP last Christmas to production services now supporting researchers nationwide. Across that journey, the team have adapted and evolved how we work. The motivation behind these changes involved a shift in how NeSI thinks about infrastructure, and is builds in new levels of flexibility, adaptability, and responsiveness. The outcomes from this investment in infrastructure are ultimately about people, their expertise, and our collective cultures.

In this talk I'll review the motivations behind NeSI's reformulated infrastructure platform, covering off the "Rebase" programme that delivered the investment. I'll look at the initial value drivers that got us started, the emerging benefits that compelled us to deliver on the investment, and the opportunities that drive us from here.

The true outcomes of NeSI's Rebase investment are yet to be seen, and will only reveal themselves over the years ahead. The early experiences and potentials are starting to emerge - I'll touch on our focus on the quality of user experiences and our guiding principles. This focus on the researcher is core to NeSI's culture and is exemplified in our work on Rebase. The focus on continuity of experience is the short-term result of our researcher-obsessed culture. As we lift our gaze to the future potentials yet to be discovered, I'll cover the longer-term horizons and aspirations we've had in mind as we designed and built a platform that is flexible, scalable, and that can adapt to the changing shape of our sector.

Ultimately this is a story about people involved in our NZ eResearch ecosystem, and specifically those shaping and rebasing our NeSI infrastructure into what we need.

ABOUT THE AUTHOR

Nick Jones is a leader of the national eScience and research High Performance Computing platform investment in New Zealand, NeSI - National eScience Infrastructure. He has been in leadership roles in the academic research sector for 20 years across eLearning and eResearch. Nick led the development of the eResearch Ecosystem Framework and sponsored a subsequent International Benchmarking Study. Nick has been an advocate for the evolution of research practice in advanced digital technologies across research data and software, computational thinking, and a user-centred and relational approach to delivering innovative experiences to researchers.

Reflecting on a Nationwide Bioinformatics Training Programme

Tyler McInnes
Genomics Aotoearoa
tyler.mcinnes@otago.ac.nz

ABSTRACT

Since 2017 Genomics Aotearoa and NeSI (New Zealand eScience Infrastructure) have collaborated to provide bioinformatics and genomics training free of charge to researchers in Aotearoa New Zealand. Following the format of The Carpentries (an internationally recognised training institution), we have run more than 100 workshops in-person and online, reaching more than 2,000 total attendees. The training programme includes a portfolio of beginner workshops covering R programming, the command line, and working on servers, as well as specialized genomics topics such as RNA-seq data analysis, long-read genome assembly, single-cell RNA-seq, and pangenome graphs.

The workshop portfolio was developed by over 30 experts from 15 institutions, including both international collaborators and local contributors who have taught internationally. The training programme is itself a valuable resource which has provided support to community initiatives (*e.g.,* IndigiData Aotearoa), fosters relationships with the private sector (*e.g.,* collaborative conferences), and offers teaching and training opportunities to Postdocs and PhD students.

This talk will showcase some of the key outputs of the training programme, highlight opportunities for researchers in Aotearoa who want to engage in bioinformatics training, and invite discussion on how we best serve the community going forward.

ABOUT THE AUTHOR

Tyler McInnes In 2022 Tyler McInnes joined Genomics Aotearoa in the position of Bioinformatics Training Coordinator where he oversees the training programme workshop portfolio, collaborating with subject matter experts in developing and instructing workshops. Prior to joining Genomics Aoteaoroa Tyler worked as a Teaching Fellow, where he developed an intervention programme to teach students critical study skills and improve pass rates. From 2019 to 2022 Tyler worked with various groups of students, delivering weekly and fortnightly workshops which supported students in learning paper content and developing skills such as: public speaking, creating and interpreting diagrams, critical assessment of questions, and many more.

Research Data Management for Research Programs – time to evolve our practices

David Medyckyj-Scott, Manaaki Whenua Landcare Research, New Zealand
Jude Channon, Centre for eResearch and Digital Innovation & Australian Research Data Commons, Australia
medyckyj-scottd@landcareresearch.co.nz
j.channon@federation.edu.au

ABSTRACT

Managing research data in multi-stakeholder, multi-project, and multi-participant research programs presents unique complexities and challenges. While substantial guidance exists on research data management (RDM) frameworks and best practices for individual projects and institutional support (e.g. the Research Data Management Framework for Institutions (2023)), there is a notable knowledge gap when it comes to the specific needs of large, integrated, interdisciplinary, and transdisciplinary research programs. This presentation will highlight the differences, challenges, and barriers observed by the authors through participation in research programs, through observing research programs, and insights from a workshop held at the Centre for eResearch and Innovation (CeRDI) in Australia in October 2024. Key issues discussed will include program data governance; cross program roles and responsibilities; data sharing practices and aligning processes across projects, partners and institutions; fit for purpose infrastructure now and into the future; creating the right data culture; and maintaining data access beyond the project lifecycle. The presentation will conclude with some initial recommendations for best practices, citing good practice examples, and identify areas where further work is needed to support effective RDM in these complex research environments.
We wish to acknowledge the valuable input of the CeRDI workshop participants in shaping this presentation.
Australian Research Data Commons. (2023) Research Data Management Framework for Institutions. DOI: 10.5281/zenodo.6392340.

ABOUT THE AUTHOR(S)

David Medyckyj-Scott is currently the Head of Data Management at Manaaki Whenua. His background is broad covering environmental science, computer science, psychology, as well a data management. Prior to his current position he held several pivotal roles overseas including being Geospatial and Research Data Manager at a national datacentre (University of Edinburgh). He led the design, implementation and operation of nationally significant geospatial and environmental data services in the UK, Europe and New Zealand, and was Technical Director at the NZ National Land Resource Centre. He has promoted and been involved in the development of open data standards.

Research Security BoF

Michael Karich & Jamie Hart
UoA, UC
Michael.karich@auckland.ac.nz, jamie.hart@canterbury.ac.nz

ABSTRACT

This Birds of a Feather talk is focused on lifting the awareness of the broad field of research and cybersecurity. In this digital age, eResearch has transformed scientific research. From the creation of large data sets, increased research collaboration, enhanced data analysis and accessibility, there is incredible potential for enhanced research impact. This digital transformation however brings security risks to the research data, researchers, their research organisations, collaborators, funders and in some cases, the country.
“What can we do as a community to enable secure research?”
This BoF aims to cover relevant topics from vulnerability management, third-party vendor management, foreign influence, to export control. We would like this BoF to be run conjointly with security and office of research staff from across the sector bridging some of the operational and administrative gaps that exist for many in the sector. We would like to highlight individual projects as well as the successes that strategic collaboration can have on more effective research administration and governance. This BoF aims to spark discussion into practical insights and actionable recommendations for our researchers, eResearch and Research office colleagues to enhance the security of eResearch initiatives and safeguard the future of scientific discovery.

ABOUT THE AUTHORS

Michael Karich is Deputy Chief Information Security Officer – Research at the University of Auckland, and the Chief Information Security Officer at NeSI. Across these roles, he strives to support and enable secure research operations across the entire data lifecycle. Before this, he held roles in HPC and cloud infrastructure, data management, and research security. With Computer Science and IT Operations qualifications, Michael pulls from a broad background to empower the secure use of new technologies in groundbreaking research. He is currently focusing on the enhancement of internationally known research capability through governance, staff enablement, and process management.

Jamie Hart is a Product Manager in the Digital Services Unit, leading the Research Experience team. Jamie’s team supports UC's digital research infrastructure and provides eResearch services to UC researchers across the research lifecycle. With a background in Law, academic publishing and the performing arts, Jamie thrives on building strong relationships with stakeholders across UC and beyond and understanding the needs of the researcher in order to drive positive change. A key focus for the Research Experience team at UC is on building the eResearch catalogue of services at UC.

Service Management by Stealth

Georgina Rae, Thomas Berger, Greg Hall, Matt Chamberlain, Nick Jones
NeSI
georgina.rae@nesi.org.nz, thomas.berger@nesi.org.nz, greg.hall@nesi.org.nz,
matt.chamberlain@nesi.org.nz, nick.jones@nesi.org.nz

ABSTRACT

As NeSI evolves to support and provide services for a growing diversity of researchers, users, and customers, the need to improve and standardise our Service Management practices has become evident. This is a trend for eResearch providers around the world, however, a full ITSM implementation is both beyond NeSI's resources and likely beyond what is needed by the research sector.
This will be a whirlwind update on how we have matured our Service Management practice over the last few years. Including,
- why service management is important,
- our hacks and approaches to build maturity of our practice over time,
- our key learnings and
- a reflection on where we have got to (so far).

ABOUT THE AUTHORS

Georgina Rae is the Science Engagement Manager at NeSI where she ensures that NeSI supports NZ's researchers and research priorities through meaningful partnerships and user-driven approaches. Before NeSI she worked in molecular biology and intellectual property.

Setting up a Data Science Core Facility from scratch

Hercules Konstantopoulos
Malaghan Institute of Medical Research
herculesk@malaghan.org.nz

ABSTRACT

In December 2023 we set out to build a Data Science core facility at the Malaghan Institute. Our goal is to provide expert data science support to our researchers and collaborators, developing robust bioinformatic pipelines, statistical models, and applications to integrate and analyse their diverse data sets. We support this goal not only by building the right capability, but also carefully evaluating what tools we build versus ones we buy. Add to this our strong focus on training, and this setup helps us meet researchers where they are and weave with their work, rather than impose novel or alien structures.
A year in we have added expertise in data science and software engineering to create a team of four technical experts in computational biology, data science, and digital technology. We have made our existing high-performance computing capability far more accessible by embracing DevOps practices. We have introduced project planning to the broader core facility and Institute. We have run training courses in programming and data presentation. And we have established a collaboration model with our lab groups that allows us to provide a predictable service, while also making time to build for the future needs of the Institute. This approach looks to various disciplines and sectors for inspiration, chiefly the world of digital technology (“big tech”). In this talk I will outline how this all comes together, what practices translate directly from other sectors, and what we had to tailor to our niche as a focussed, independent Institute in our relatively isolated island nation.

ABOUT THE AUTHOR

Hercules Konstantopoulos - Growing up by the mountains of Northern Greece, Hercules Konstantopoulos developed a fascination with the night sky and all its intrigue. After a career as a researcher in astrophysics that spanned ten years and four continents, he became drawn to addressing a greater variety of data-related problems. Data science ensued with work on sustainability, renewable energy, enterprise software, and now medical research. His work focuses on converting information into strategy, and on crafting useful tools, apps, and visuals.

Shaping our conferences. eResearch conferences in Aus and NZ moving forward.

Luc Betbeder, Nick Jones, Jana Makar
AeRO, NeSI
luc@unsw.edu.au n.jones@auckland.ac.nz jana.makar@nesi.org.nz

ABSTRACT

Our eResearch conferences in Australia and NZ have been running for 15 years. We propose to run a BOF session to discuss some conference lessons and identify opportunities to align as we move forward. The main BOF topics would include:

- Review. What are the big conference lessons since COVID? Let's discuss in-person conferences vs online and the dreaded hybrid. Do we expect our in-person eReserach conferences to continue? We have had a successful eRA with 600 attending in Melbourne.
- Review. Is "eResearch" still a helpful term for us? We have a good track-record and brand-identity but it is becoming a less common term with our CIO colleagues and vendor partners. If we want easier conference approvals should we change the name?
- Going forward. Is one of the roles of these two "conferences" to be one of the key pillars in professional development and workforce sustainability? If yes what else could we offer through our conference programs.
- Going forward. Is there space for an online only conference in April/May/June to re-use some of our in-person content? Would this be good for equity and access to conference content or not really worth the extra effort.
- Going forward. Are there any other opportunities to align our conferences and reach out to wider Australasia?
- Questions arising from our online document and discussion.

The org and program committees of eRA 24 and eRNZ 25 will be emailed these questions and links to doc to kick-start our discussions.
The BOF will try to keep all participants together and move sequentially through the questions above. A summary of findings will be shared with the committees.
Working groups may be formed to follow up items.

ABOUT THE AUTHOR

Luc Betbeder is President of AeRo (Australasian eResearch Organisations) and co-chair of the eResearch Australasian peak Annual Conference. He is an Adjunct Senior Lecturer in UNSW Faculty of Medicine Centre for Big Data and has been a Visiting Scientist with the Data61 Visual Analytics Team in Australia's National Science Agency CSIRO. He has held director-level roles for 15 years in higher-ed ICT and eResearch. Luc is currently the Executive Director Research Technology Services at UNSW, a shared services function that he established which is responsible for providing Research Computing and Research Data to 5000 academic staff. Recently, Luc has been working with colleagues to regularly count how much Research Data there is in Australia and New Zealand.

Streamlining a National Flood Assessment using a Flow Scheduler

Harang, A.,[1] Lane, E. M.,[1] Bosserelle, C.,[1] Dean, S.,[2] Cattoën, C.,[1] Pearson, R.,[1] Carey-Smith, T.,[2] Srinivasan, R.,[2] Shiona, H.,[1] Wilkins, M.,[1] Smart, G.,[1] Oliver, H.,[2] Lin, Y.[2]
[1] NIWA Christchurch
[2] NIWA Wellington

ABSTRACT

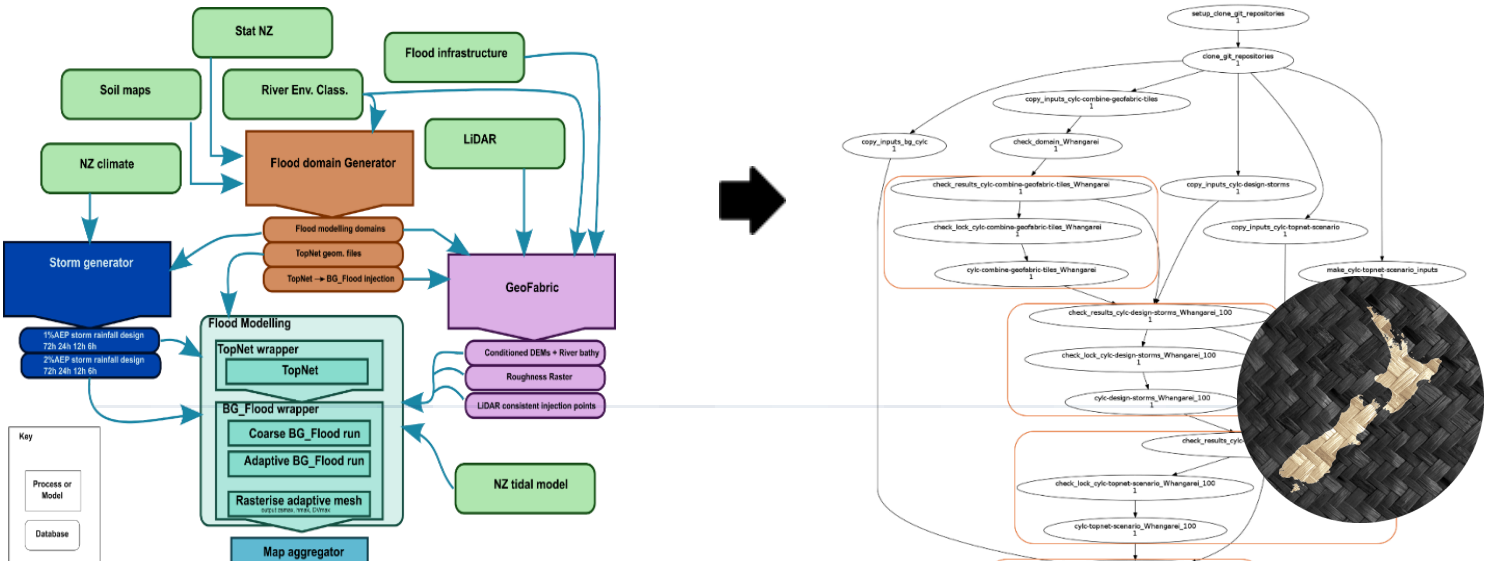
Flooding is one of the costliest natural hazards in Aotearoa and climate change is increasing the frequency and severity of floods. We need a national-scale flood hazard and risk assessment under the current and future climate to plan for this. The Endeavour project “Mā te haumaru ō nga puna wai ō Rākaihautū ka ora mo ake tonu: Increasing flood resilience across Aotearoa” (http://niwa.co.nz/flood-resilience) aims to produce a consistent and automated method to generate flood maps nationwide. These flood maps will then be used for risk assessment, inform on societal vulnerability and identify solutions to reduce and adapt to flood risk.

A cylc workflow (https://cylc.github.io/) was built to generate these flood maps. Cylc was initially designed for weather forecasting and scheduling jobs by job dependencies and time increments. In this workflow, cylc is principally used as a central tool to link the different models and loop through all the computational domains (based on flood plains and their catchments) and scenarios. As the flood modelling needs to integrate climate science, rainfall statistics, hydrology, hydrodynamics, and geospatial data for each domain, a modular workflow was built. Each of these main tasks spins up a cylc sub-workflow (see Figure). Each subworkflow can also be run independently for testing and validation. This allows the high-level workflow to be simple, accessible and usable by a larger group of researchers. Through this modularisation, each sub-workflow can be designed specifically for its scientific domain. They then communicate smoothly through the main workflow. Each workflow is subdivided into tasks that are submitted independently to the workload manager (Slurm on NeSI). This functionality creates flexibility and allows the appropriate machine, computational environment and computing resources to be allocated for each job. For example, this workflow uses GPUs and parallel CPUs, as well as different computing languages such as C++, CUDA, Fortran90, Python, R, Julia and batch.

The workflow has been used for the generation of the first set of flood maps. It iterated over 248 computational domains. For this first test, two flood maps corresponding to different scenarios have been generated for each domain. This will be increased next run with the inclusion of climate projection scenarios.

This set-up, version controlled using git and deployed on NeSI, provides a robust method for collaborative code development, reproducibility and flexibility.

The author(s) wish to acknowledge the use of New Zealand eScience Infrastructure (NeSI) high performance computing facilities, consulting support and/or training services as part of this research.



ABOUT THE AUTHOR

Alice Harang is a hydrodynamic modeler part of the Natural Hazard group at NIWA. She has experience in computational fluid dynamic with modelling of small-scale mixing mechanism to large scale coastal flows or inundation scenarios. She grew her experience in flood modelling particularly through the ENDEAVOUR project “Mā te haumaru ō nga puna wai: Increasing flood resilience across Aotearoa” where she co-developed a workflow to create flood maps automatically across all New Zealand. She also was central in the creation of flood maps following TC Gabrielle for the recovery process, working with local councils, consultants, and university students. Finally, she is co-developing the 2D numerical model BG_Flood, a shallow water solver, producing to model inundation from compound sources such as tsunami, rain, storm surge, or river discharge.

Teaching HPC Bioinformatics to Researchers: The University of Melbourne Experience

Lev Lafayette, Daniel Tosello
University of Melbourne
lev.lafayette@unimelb.edu.au d.tosello@unimelb.edu.au

ABSTRACT

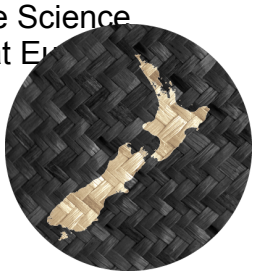
The University of Melbourne is regularly assessed as one of the world's leading universities and much of this is a result of research strengths in the health sciences with several significant affiliated institutions. In the computational space, developments in the health sciences are heavily dependent on the broad and interdisciplinary discipline of bioinformatics. The use of bioinformatics applications, especially those with complex workflows or large datasets, is most effectively and efficiently conducted in a high performance computing (HPC environment). However, as well documentated in previous publications and presentations, many researchers are unfamiliar with HPC environments, often painfully evident in bioinformatics software. Basic education of these environments often becomes the responsibility of HPC centres themselves.

At the University of Melbourne, there is a general-purpose HPC system, "Spartan", along with some smaller systems housed in affiliated institutions. Spartan's history starts with an innovative and cost-effective architecture and which now is a world's leading system with Top500 certification. Since Spartan's introduction a key feature of the enviroment has been an extensive user education programme, which is strongly correlated with very high levels of job submission. Specialist discipline-specific training workshops are also an on-going feature of the programme, including bioinformatics, with a training approach that is sensitive to the cultural backgrounds of the user community.

This presentation outlines the HPC bioinformatics teaching content and experience at the University of Melbourne. It is based on the recognition that HPC systems will continue to develop and incorporate new technologies that improve performance as data complexity and size increases. In fact, the evidence is that researchers increasingly require access to HPC systems and knowledge of how to operate in this growing environment. The University of Melbourne programme, covering operating system knowledge, application use, environment modules, job submission, web-based interactive environments, and parallel processing skills, uses a scaffolding approach starting from the very basics, to a distributed training programme with the "Spartan Champions" project. The successes of this programme is suitable for adoption by other institutions that also wish to educate researchers in HPC use, not just in bioinformatics, but other disciplinary areas.

ABOUT THE AUTHORS

Lev Lafayette is a Senior HPC DevOps engineer and the HPC Solutions Team Leader at the University of Melbourne, where he has worked for the past ten years. Prior to that, he worked at the Victorian Partnership for Advanced Computing in a similar role. Over the past twelve years, he has taught thousands of researchers across more than twenty Australian Universities and government agencies. An incorrigible collector of degrees and with a wide range of interests, Lev has just completed his fourth master's (in Climate Change Science and Policy) and has eight degrees to his name. He has just started a doctorate at Eu University.



Daniel Tosello is a HPC DevOps engineer at the University of Melbourne, where he has worked for the past ten years. Prior to that he worked as a software engineer at La Trobe University and for ICT client support and also as an IT support officer at the AARLIN Consortium. Daniel has recently presented at eResearchAustralasia with representatives of Victorian health organisations on tailoring HPC for medical research. He also has interests in 3D printing and CNC pipelines.

The “s” in “eResearch” Stands for Security

Sean Matheny, Tom Laurenson, Michael Karich
Waipapa Taumata Rau, University of Auckland
s.matheny@auckland.ac.nz, tom.laurenson@auckland.ac.nz, michael.karich@auckland.ac.nz

ABSTRACT

Compared to other computing and data domains, eResearch has historically overlooked general information security processes. Initially, the focus was on open collaboration, which contrasts with the restrictive boundaries that information security seeks to implement. Additionally, eResearch involves complex, bespoke, and temporary workflows, deterring the adoption of corporate one-size-fits-all solutions. Furthermore, a general perception of low risk, combined with funding and resource constraints, has led to reduced uptake of cybersecurity practices.

However, this is rapidly changing. Both researchers and IT departments are being compelled to quickly enhance their security posture. The proliferation and increasing sophistication of cyber-attacks have introduced not only financial and operational risks but also reputational risks for both institutions and researchers. Moreover, research data increasingly includes sensitive information, making it attractive to malicious actors.

The positive development is that academic institutions are beginning to recognize the importance of security and the unique challenges within the eResearch space. We are collaborating with researchers to find balanced solutions that maintain the open collaborative model while incorporating layered security controls. This presentation will focus on the current challenges and potential solutions, with particular emphasis on lessons learned from operating the Auckland node of ARDC’s Nectar Research Cloud.

ABOUT THE AUTHORS

Sean Matheny began his career working in the physics department of the largest radiation oncology treatment provider in the U.S. There, he helped to improve QA, dosimetry, and treatment planning and delivery via technology, machine learning, and automation. For the last 8 years, he’s primarily worked in roles that have enabled research cloud platforms and HPC (high performance computing) in both the Centre for eResearch (University of Auckland), and NeSI. Currently, he helps operate the Auckland node of the Nectar Research Cloud, at Centre for eResearch.

Tom Laurenson is a Senior Security Engineer in the Centre of eResearch Platform team at the University of Auckland. His primary responsibility is improving the security posture of research at UoA, from a technical standpoint. Tom recently returned to academia, after various technical positions in the industry as a security engineer and pentester. Before that, he was a lecturer at the Otago Polytechnic after receiving a PhD from the University of Otago. Tom is passionate about programming, breaking stuff, security, and finding the harmonious balance of risk, security and getting things done.

Michael Karich is Deputy Chief Information Security Officer – Research at the University of Auckland, and the Chief Information Security Officer at NeSI. Across these roles, he strives to support and enable secure research operations across the entire data lifecycle. Before this, he held roles in HPC and cloud infrastructure, data management, and research security. With Computer Science and IT Operations qualifications, Michael pulls from a broad background to empower the secure use of new technologies in groundbreaking research. He is currently focusing on the enhancement of internationally known research capability through governance, staff enablement, and process management.

The Case for Open Source in Aotearoa *(a continuing conversation)*

Richard Littauer, Jonah Duckles
Te Herenga Waka Victoria University of Wellington, Organizational Mycology
richard.littauer@vuw.ac.nz, jonah@orgmycology.com

ABSTRACT

Last year, we discussed scientific open source in Aotearoa New Zealand, looking at projects that we admired, then obstacles to releasing work more openly, what benefits and support would be necessary for those projects to thrive. Then, we discussed what keeps the open source ecosystem in Aotearoa New Zealand from reaching a critical mass, and why there isn't more open source. We want to follow that this year with a discussion on ways that we to ensure funding and buy-in from others, particularly under administrations that are more focused on corporate interests.

We hope to have an explicit discussion on how we can ensure that open source is not an unfunded mandate for researchers, but instead a funding-multiplier, enabling more pathways to corporate collaboration, providing a catalyst for international collaborative investment – particularly within scientific funding ecosystems like government initiatives like EU Horizons or NSF and NIST, or private funders like CZI, and a way for researchers to build out their skills. The potential impacts of open source have not been fully explored from an NZ perspective. In this BoF, we hope to both allow local participants and leaders to share use-cases from their perspective, but also to talk about strategies that would enable them to pitch up and laterally. Part of this will also involve listing the open source tooling that we already use, and which is already showing the professional and institutional advantages of developing in the open.

The organizers plan to bring some examples of how to use open source to collaborative effectively and sustainably, based upon their experience with funders and communities abroad. However, we hope that others will come with their own perspectives, too, so that we can collaboratively begin work on a game plan for how to talk about Open Source within an the Aotearoa New Zealand environment.

ABOUT THE AUTHORS

Richard Littauer is a PhD student in Computer Science at Te Herenga Waka Victoria University of Wellington in Pōneke, Aotearoa New Zealand. His primary focus is understanding ecology and bird populations using computational modeling. He is also the Interim Executive Director for the GNOME Foundation, an organizer for SustainOSS, and one of the two organizers of CURIOSS, the community for university and research institution open source program offices. He has been interested and involved in open source communities for decades. He is also on the committee for PythonNZ, and an ECR for the Royal Society Te Apārangi. You can read more about his projects at his website, https://burntfen.com, and follow him on http://richard.social.

Jonah Duckles has built a diverse career as an organizational leader across various sectors. At Organizational Mycology he and the team explore the intersection of people, technology, and building organizations that can maintain and sustain their impact.

Keynote: The Digital Chemistry Revolution—A Synergy of HPC, AI, and Quantum Mechanics

Dr Giuseppe Barca
Associate Professor, The University of Melbourne

ABSTRACT

Scientific progress in fields such as drug discovery, sustainable manufacturing, and materials science increasingly depends on our ability to precisely simulate matter at large atomistic scales. Current simulation methods, however, remain slow and prone to inaccuracies, forcing heavy reliance on physical experimentation that introduces delays, increases costs, and carries inherent limitations.

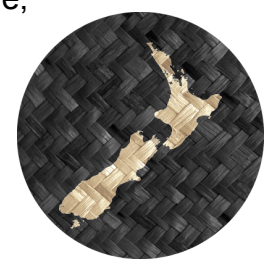
We introduce a groundbreaking digital chemistry paradigm that addresses these challenges by integrating many-GPU high-performance computing, machine learning, and computational quantum chemistry. By leveraging novel algorithms and software optimizations, we overcome the steep computational cost and efficiency limitations that have historically hindered the use of quantum chemistry at large molecular scales and useful speed. Awarded the 2024 Gordon Bell Prize, our approach enables, for the first time, quantum-level simulations of complex bioscale molecular systems with unprecedented accuracy, approaching that of physical experiments

Through exascale supercomputing, automated workflows, and machine learning models trained on quantum-accurate data, we streamline end-to-end simulation and design processes. These intelligent workflows reduce dependency on laboratory experiments, offering a fully automatable, cost-effective, adaptive, and highly precise solution for investigating complex molecular phenomena.

The result is a powerful platform that accelerates discoveries and transforms research across chemistry, biology, and engineering. This keynote will detail how to practically implement and leverage this paradigm, providing scientists and industries with a clear roadmap for harnessing HPC, AI, and quantum mechanics to foster groundbreaking innovations and drive tangible real-world impact.

ABOUT THE AUTHOR

Giuseppe Barca specialises in interdisciplinary research at the intersection of high-performance computing (HPC), artificial intelligence (AI), and digital chemistry at the University of Melbourne, Australia.He has led collaborations with top universities, supercomputing centres, and companies across Australia and the US. From 2018 to 2023, he was heavily involved in Australia's contribution to the US Department of Energy's Exascale Computing Project (ECP) for GAMESS, creating advanced computational chemistry software for systems like Frontier and Aurora. He led a project funded by Australia's Pawsey Centre for the Extreme-scale Readiness (PaCER) program, enhancing molecular modelling capabilities. His research advances digital technologies with HPC and machine learning to accelerate chemical research, enabling high-speed, precise predictions in molecular systems. His group's AI-powered software aids in designing new therapeutics, catalysts, and materials for academia and industry. From 2020 to 2022, his team set four world records in quantum chemical modelling, utilising over 27,000 GPUs on Summit and 37,000 GPUs on Frontier. In 2023, he co-founded QDX Technologies, a deep-tech company focused on drug discovery via quantum chemistry and AI, operating in Singapore, Melbourne, and Canberra.



In December 2024, Giuseppe and his research team were awarded the ACM Gordon Bell Prize in HPC, and earlier this year Giuseppe was recognised as one of Australia’s Top 250 Researchers for his impactful, highly cited work.

The Impact of AI Technologies on eResearch and Data Centre Design

Andrew Kirker,
CDC Data Centres NZ Limited
andrew.kirker@cdc.com, richard.prowse@cdc.com

ABSTRACT

Artificial Intelligence (AI) is fundamentally transforming the way we store, process, and analyse data, bringing about significant changes in cloud computing and data centre management. This evolution is creating unprecedented opportunities for organisations to enhance efficiency and innovation, while also presenting new challenges in areas such as security, privacy, ethics, and governance.

AI technologies are enabling more intelligent and automated management of cloud and data centre resources. Machine learning algorithms can predict usage patterns, optimize resource allocation, and reduce operational costs. AI-driven analytics provide deeper insights into data, facilitating better decision-making and strategic planning. Moreover, AI enhances security by detecting and responding to threats in real-time, thus safeguarding critical data assets.

However, the integration of AI into cloud and data centre strategies is not without its challenges. Organisations must address complex issues related to data privacy and security, ensuring that AI systems are transparent, fair, and compliant with regulatory standards. Ethical considerations are paramount, as biases in AI algorithms can lead to unintended consequences. Robust governance frameworks are essential to manage these risks and ensure responsible AI deployment.

This presentation will delve into the transformative impact of AI on cloud and data centres. We will explore practical applications, such as automated resource management and advanced threat detection, and discuss the ethical and governance challenges that accompany AI integration. Attendees will gain a understanding of how to leverage AI to optimize their cloud and data centre strategies while navigating the associated complexities.

ABOUT THE AUTHOR

Andrew Kirker, Managing Director for New Zealand and Hyperscale ANZ at CDC, joined the company in 2019. With over 25 years' experience delivering critical systems to data centres, Andrew has overseen the development and growth of data centre businesses across Australia and New Zealand for leading industry brands. Previously, at American Power Conversion (APC) and later Schneider Electric, he played a key role in establishing and transforming data centre infrastructure. At CDC, he has been pivotal in expanding operations to New Zealand and leads customer growth and strategy. Andrew is also a respected industry speaker on market trends and technology infrastructure.

The many talents of leadership in open science & open source projects

Jonah Duckles, Dan Sholler, Beth Duckles
Organizational Mycology
jonah@orgmycology.com, dan@orgmycology.com, beth@orgmycology.com

ABSTRACT

Recently in 2024, Organizational Mycology has run six community calls for the Chan Zuckerberg Initiative's Essential Open Source and Open Science programs. In these community calls we convened a 1-hour discussion and created a community resource based on each discussion. In this presentation we'll give a summary of these resources, and the hidden work, skills and knowledge that project leaders wish they had known. Taking input from project leaders who are new to leading emerging projects to veteran leaders from major and lasting open source projects, these resources draw from the collective wisdom of other open source / open science leaders.
Across each of these resources, we'll talk about high-level project considerations, leading people without formal authority over them, measurement and assessment, sustainability, financial models and managing project documentation efforts. In a short 15-minutes, participants will come away with a keen understanding of how these resources could be useful to them in their own careers as they step into leadership positions of all kinds, open source or otherwise. To close, we'll revisit last year's eResearch BoF discussion on the challenges and opportunities in creating open source software projects in the New Zealand science ecosystem.

ABOUT THE AUTHOR

Jonah Duckles has built a diverse career as an organizational leader across various sectors. At Organizational Mycology he and the team explore the intersection of people, technology, and building organizations that can maintain and sustain their impact.

The State of Open Data, Data Citation and Reuse

Anthony Dona PhD
Digital Science
a.dona@digital-science.com

ABSTRACT

One of the pillars of Open Science is the publication of research data. Firstly, open research data supports the reproducibility and scrutiny of research. Secondly, re-use of the data by others can make research more efficient. Through the State of Open Data[1] report, we can see the transition in attitudes of academics, as well as regional policies towards the sharing of research data.

We are trying to measure the adoption of these ideas in the scientific community by counting datasets and their citations in publications using Dimensions data on Google BigQuery (GBQ).

We show that the number of cited datasets and citing publications has increased massively however there remains a very small subset of the registered datasets. We analyse the numbers by repository and subject and raise several questions around citation policy as well as future avenues of research.

We also delve further into the state of data sharing in New Zealand and compare the attitude towards data sharing, the methods for sharing, rates of citation and reuse compared to other regions across the globe.

[1] Hahnel, Mark; Smith, Graham; Schoenenberger, Henning; Scaplehorn, Niki; Day, Laura (2023). The State of Open Data 2023. Digital Science. Report. https://doi.org/10.6084/m9.figshare.24428194.v1

ABOUT THE AUTHOR

Anthony Dona is a Senior Director at Digital Science, using his role to drive his passion for enabling novel research outcomes. He works specifically with government and funding organisations across the Asia-Pacific region to aid data-driven decisions related to research funding and assessment. Anthony previously gained a PhD from the University of Sydney in Biochemistry and has worked as a postdoctoral researcher at both the University of Sydney and Imperial College London.

A BoF in two parts: Maturing research data management (RDM) and enabling good data management planning practices

Yvette Wharton[1], Laura Armstrong[1], Sarah Hopkins[1], Mark Gahegan[1], Richard Hartshorn[2] David Medyckyj-Scott[3], Claire Rye[4]

[1] Waipapa Taumata Rau, University of Auckland, [2] Te Whare Wānanga o Waitaha University of Canterbury, [3] Manaaki Whenua Landcare Research, [4] New Zealand eScience Infrastructure
y.wharton@auckland.ac.nz, l.armstrong@auckland.ac.nz, s.hopkins@auckland.ac.nz, m.gahegan@auckland.ac.nz, richard.hartshorn@cantebury.ac.nz, Medyckyj-ScottD@landcareresearch.co.nz, claire.rye@nesi.org.nz

ABSTRACT

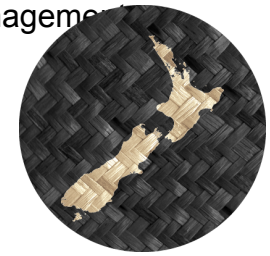
Enabling good research data management (RDM) practice is increasingly important given the intersection of data, ethics, security, new technologies and reproducible research. To do this, universities and other research organisations undertake maturity assessments and develop roadmaps to provide efficient and practical training, connected tools for data management planning (DMPing), infrastructure, and support services. These enable researchers to undertake responsible research data management and stewardship practices that align with policy and regulatory requirements whilst supporting organisational requirements.
In the first BoF session, we will facilitate group discussions to gather and share stories on your RDM journey, drivers, priorities, barriers, benefits from different viewpoints and RDM experiences.
We will come back in the second session to support an open conversation exploring the issues and opportunities identified in the first BoF and the activities, solutions, and practices that have worked for your institution and researchers.
These two BoF sessions are aimed at those providing research infrastructure services and support in data management planning across the research sector, including researchers, research data stewards, data managers, research support staff, institutional administrators, enterprise architects, eResearch directors, and academic engagement or research librarians.

ABOUT THE AUTHORS

Yvette Wharton works at the Centre for eResearch, Waipapa Taumata Rau, University of Auckland. She is the Technical Business Lead for the Research Data Management Programme, working on the Secure Research Environment and machine-actionable Data Management Planning initiatives. http://orcid.org/0000-0002-6689-8840

Laura Armstrong is the eResearch Engagement Lead for Waipapa Taumata Rau, University of Auckland. She collaborates to engage with the research community to raise awareness and use of modern technologies and tools to advance research.
http://orcid.org/0000-0003-2370-3924

Sarah Hopkins is an eResearch Engagement Specialist in the Centre for eResearch, Waipapa Taumata Rau, University of Auckland. She is passionate about using her research experience to support researchers to navigate the evolving requirements for research data management through awareness, engagement and supporting robust data management planning. https://orcid.org/0000-0002-4705-5362



Mark Gahegan is the founding director of the Centre for e-Research and Professor in the School of Computer Science at the University of Auckland, New Zealand. Mark has wide interests at the intersection of Computer Science, Machine Learning, eScience and GIS Science. This is because he can never sit still for too long and is easily distracted. He is kind of scruffy-looking. He also directs the 'Beyond Prediction...' national Data Science programme, focused on applying emerging Machine Learning, eScience and GenAI methods to problems in Ecology, genomics and infectious diseases.

Richard Hartshorn is a Professor of Chemistry in the School of Physical and Chemical Sciences at the University of Canterbury. He is heavily involved in data initiatives within the International Union of Pure and Applied Chemistry (Secretary General 2016-2023) and in CODATA (elected Vice President 2023). He leads the Aotearoa New Zealand Committee on Data in Research. Richard Hartshorn (0000-0002-6737-6200) - ORCID

David Medyckyj-Scott is currently the Head of Data Management at the Manaaki Whenua. His background is broad covering GIS, psychology, computer science as well as geospatial and environmental data management. Prior to his current position he held several pivotal roles including Geospatial and Research Data Manager at a UK national data centre. There and in New Zealand he led the design, implementation and operation of nationally significant geospatial and environmental data services, and was Technical Director at the NZ National Land Resource Centre.

Claire Rye is a Product Manager at New Zealand eScience Infrastructure (NeSI) based out of the University of Auckland. She is responsible for the National Data Transfer Service and works across the Aotearoa Genomics Data Repository and Rakeiora Pathfinder projects and looking at research data management and data lifecycle more generally across NeSI, and is a co-chair for the World Data System Early Career Researchers network. Claire holds a PhD in organic chemistry and has spent the last 11 years working in the UK in a variety of research settings.

What Really Drives Bioinformatic Tool Accuracy?

Paul P Gardner
University of Otago, Dunedin

ABSTRACT

In bioinformatics, multiple tools often exist for the same task, yet their accuracy can vary significantly. I have led several projects evaluating the accuracy of bioinformatic software. One of which is focused on detecting protein-coding sequences in nucleotide sequences, which are commonly used in long non-coding RNA pipelines [1]. Benchmarks like this prompted my team to explore what factors are linked to software accuracy [2]. We found that citation-based metrics (H-index, impact factors, citations) had no correlation with accuracy. Instead, indicators of long-term software support, such as GitHub activity, were strongly associated with better performance. This suggests that sustained support for bioinformatics tools is more beneficial than pursuing citation-based reputation. I will conclude with a grudge-based investigation into the link between academic department affiliation and software accuracy [3]. These studies highlight the crucial role of continuous development and interdisciplinary collaboration in producing reliable bioinformatic software.

1. Champion DJ, Chen TH, Thomson S, Black MA, Gardner PP (2024) Flawed machine-learning
confounds coding sequence annotation. bioRxiv.
https://doi.org/10.1101/2024.05.16.594598
2. Gardner PP et al. (2022) Sustained software development, not number of citations or journal choice, is
indicative of accurate bioinformatic software.
https://doi.org/10.1186/s13059-022-02625-x
3. Gardner PP (2024) A Bioinformatician, Computer Scientist, and Geneticist lead bioinformatic tool
development - which one is better? https://doi.org/10.1101/2024.08.25.609622

Why we ditched shiny in favour of react

Richard Dean
Institute of Environmental Science and Research (ESR)
richard.dean@esr.cri.nz

ABSTRACT

When is it time to ditch the known tech stack and embrace something new? For ESR's data science team, the pandemic forced us to develop advanced geospatial analysis tools at pace. We pushed the limits of the likes of R shiny, and while toolkits like {golem} helped us to build production grade dashboards, we found that our ambitions were limited by the frameworks.
More recently, we've settled on the React framework for advanced applications. In this presentation, Richard will present several use cases which were simplified through the adaptation of React.
Examples will include:

- deploying machine learning to web browsers, and benchmarking performance
- the integration of large language models into ESR's digital twin
- new ways to visualise agent based models using Cesium
- recharts and d3 for interactive graphs and process visualisation
- how we're using libraries such as immer and zustand to maintain state
- how we're using react, express and sequelize to build standalone APIs
- how we build and deploy react apps

To conclude, we will argue that React offers a simple and flexible option to develop and deploy advanced AI powered data visualisation tools – and, as such, it deserves a higher profile in the eResearch community.

ABOUT THE AUTHOR

Richard Dean is a senior data scientist in ESR's core data science team. He has over 20 years' experience working with health, forensic and environmental data sets, specialising in the development of novel data tools and visualisation techniques. He currently leads a research programme looking at the use of computer vision for rapid diagnostics and is responsible for the development of ESR's digital twin user interface.

Keynote: Becoming a Researcher - Building Skills, Community, and Shaping the Digital Future

Dr Kari Jordan
The Carpentries
kariljordan@carpentries.org

ABSTRACT

Who shapes the digitally skilled workforce—not just for today, but for the future? In this keynote, I will reflect on my own journey into research and the pivotal role that community and collaboration play in digital upskilling. Drawing insights from The Carpentries' global, volunteer-driven training ecosystem, we'll examine the evolving landscape of research skills development—what's changing, what must endure, and how we sustain progress. Together, we'll navigate the challenges and opportunities that lie ahead, exploring practical strategies for fostering an inclusive and resilient digital research community. You'll leave with fresh perspectives, new connections, and the confidence to actively shape a resilient and empowered digital future—together.

ABOUT THE AUTHOR

Dr. Kari L. Jordan is a visionary leader and advocate in data science education and community building. As the Executive Director for The Carpentries, an internationally recognized organization dedicated to teaching researchers foundational coding and data science skills, Dr. Jordan has played a pivotal role in shaping the landscape of data literacy and empowering individuals worldwide.

With a Ph.D. in Engineering Education focusing on the sense of belonging for people of color in STEM, Dr. Jordan combines technical expertise with a deep commitment to fostering inclusive learning environments. Her work centers on making computational skills accessible to all, bridging gaps in technical knowledge, and creating opportunities for underrepresented groups in data science. Under her leadership, The Carpentries has grown its global footprint, forming partnerships with academic, government, and industry leaders. Dr. Jordan's strategic vision and dedication to equity have driven initiatives that amplify diversity and representation in the data science community, ensuring individuals from all backgrounds can thrive.

In addition to her executive role, Dr. Jordan is a sought-after speaker and mentor, sharing her expertise and insights at international conferences and workshops. Her collaborative and empathetic approach has inspired countless individuals to pursue careers in data science and empowered them to make a positive impact in their respective fields.

Dr. Kari L. Jordan's unwavering commitment to data science education, exceptional leadership skills, and passion for equity and inclusion have positioned her as a driving force in shaping the future of data literacy. Her contributions to The Carpentries and the broader data science community continue transforming how individuals acquire and apply critical computational skills, ultimately fostering a more inclusive and empowered society.

Bridging Borders: Your Gateway to Global Research Data Collaboration

Trish Radotic
RDA and ARDC
trish.radotic@ardc.edu.au

ABSTRACT

Discover the Research Data Alliance (RDA), the unique global community fostering open science and advancing research data collaboration. This session highlights RDA's impact in connecting researchers, data professionals, and research infrastructure experts across borders to tackle some of the world's most complex challenges. Learn how you can get involved, contribute to RDA outputs and recommendations, and leverage the global RDA community ecosystem to accelerate research data solutions, drive innovation, and support open science initiatives.

ABOUT THE AUTHOR

Trish is the Research Data Alliance (RDA) Regional Community Manager (Oceania and East Asia) at Australian Research Data Commons (ARDC). With over 20 years' experience in IT, marketing and international business, Trish supports the RDA Secretariat's mission of building the social and technical bridges to enable the open sharing and reuse of research data across technologies, disciplines and countries. Trish is passionate about advancing global collaboration through connection, networking and community building.

Beyond the Data: How AI and Metadata Unlock True Value from Content to Research

Alex Timbs
Dell Technologies
alex.timbs@sempre.co.nz

ABSTRACT

In both media production and research, the challenge isn't just managing vast amounts of data—it's making sense of it. Content creators and researchers alike grapple with data deluge, discoverability, classification, visualization, and workflow bottlenecks. This session explores the striking parallels between media pipelines and research workflows, highlighting how AI and metadata-driven automation transform raw information into meaningful insights. From streamlining VFX production to accelerating scientific discovery, we'll examine how innovations in media pipelines might inform and enhance research strategies

ABOUT THE AUTHOR

Alex Timbs is the Business Development and Alliances Manager at M&E Dell Technologies. Joining Dell Technologies five years ago, Alex is focused on expanding his expertise in managing, storing and orchestrating data, taking on the role of Business Development and Alliances Manager within the Media and Entertainment vertical. Alex brings over 15 years of experience from Netflix Animation, formerly known as Animal Logic, where he spearheaded technology teams across Sydney, Vancouver, and Los Angeles. During his tenure, he was instrumental in delivering high-density containerised data centres, overseeing the expansion of virtualized infrastructure, and adeptly navigating the ebb and flow of resource demands throughout the production of numerous award-winning animated and visual effects features. With a sincere belief in the symbiosis of technology and art, Alex advocates that one cannot thrive without the other in today's digital landscape. He is particularly passionate about the value of AI in content creation, seeing it as a means to accelerate ideation and iteration. His approach is rooted in the conviction that understanding and articulating the "Why" is paramount to unlocking true business value. His commitment to cultivating trust as the foundation of any successful business relationship underscores this belief.
Alex understands the critical role of metadata-driven automation in today's content-rich environments, championing its ability to streamline creation workflows, enhance discoverability, and drive efficiency in content management and delivery processes. This enhanced focus ensures that Alex remains at the forefront of the industry, pushing the boundaries of what can be achieved at the intersection of technology, business, and art.