

AOTEAROA  
eRangahau  
Research<sup>NZ</sup>

2024

7-9 February 2024  
15th annual conference  
**WELLINGTON**

*Navigating an Evolving eResearch Landscape*  
*He Moana Pukepuke e Ekengia e te Waka*

PLATINUM



GOLD



SILVER



BRONZE



NETWORKING DRINKS



LUNCH SPONSOR



INTERNATIONAL  
DATA WEEK **2025**  
— BRISBANE AUSTRALIA  
13-16 OCTOBER 2025 —



Learn more and sign-up for updates - [bit.ly/ardc-idw25](https://bit.ly/ardc-idw25)

Join the international festival of data in 2025 with a global community of researchers, data professionals, policy makers and more to celebrate and advance the role of data in science and society.

International Data Week 2025 is a global event bringing together researchers, data professionals, policy makers and more to celebrate and advance the role of data in science and society.

The event will showcase how data can be used to address global challenges, foster collaboration and inclusion, and improve research quality and impact.

*This event is for:*

- Data scientists
- Data professionals
- Researchers
- Industry leaders
- Entrepreneurs
- Policy makers
- Data stewards

# “Magically” creating and updating research articles from experiments

Mark Gahegan<sup>1</sup>, Gus Ellerm<sup>2</sup>, Ben Adams<sup>2</sup>

<sup>1</sup> The University of Auckland; <sup>2</sup> The University of Canterbury

[m.gahegan@auckland.ac.nz](mailto:m.gahegan@auckland.ac.nz), [gus.ellerm@pg.canterbury.ac.nz](mailto:gus.ellerm@pg.canterbury.ac.nz), [benjamin.adams@canterbury.ac.nz](mailto:benjamin.adams@canterbury.ac.nz)

## ABSTRACT

When all steps in the process of analysis—from data discovery and integration, to method development and revision, to application in a given domain, to final publication—are made reproducible and transparent, we close the gap between *doing* research and *communicating* research. However, it is estimated that up to 80% of published research cannot be reproduced or reused because of missing, incorrect or inaccessible information. This is shocking, bearing in mind that much research is now conducted almost entirely *in-silico*. In theory we should be able to capture and manage all the relevant details because they all exist somewhere on our servers!

Traditional research journals are fossilized objects that often contain errors, ambiguities and are completely disconnected from the originating scientific process. They are often out of date before they appear in print, or become so quickly when new datasets or better methods are developed. By leveraging integrations, not before possible, between maturing eScience technologies, we are able to approach publication in more innovative ways.

What if research articles could be authored differently? What if writing the article *caused* the experiments to be conducted? What if conducting the experiments *caused* the article to be written? Both of these approaches are fast becoming possible. Our talk here considers the second of these questions: automatically creating and updating research articles from experiments.

Using *Globus* and *Gladier* (<https://github.com/globus-gladier/gladier>) in collaboration with Argonne Labs, we have created a workflow system that supports reactive, dependency-based computations to facilitate truth maintenance when changes are made to the data or methods. We have containerised these workflows using the Research Object Crate specification (*RO-Crate*: <https://www.researchobject.org/ro-crate/>). A series of nested *RO-Crates* contain descriptions of the experiment at progressive levels of abstraction/detail. The highest *RO-Crate* container is not the entire workflow, but the research article itself. The article is thus fundamentally *bound to* the experiment and in large part created from it. The article then becomes the final abstraction of the experiment into the familiar form of a published paper.

The talk will address two questions:

1. How can we link together all of the steps in an analysis workflow so that an experiment remains *live*, that is: reactive to changes in both data and methods? If we can achieve this, it becomes possible to write papers that can update themselves when better data and methods become available. Think for example of research that describes the state of a pandemic, or the impacts of climate change on coastal communities. Such papers could in theory stay up-to-date and relevant even after publication because they are dynamically created from these changing resources.
2. To what extent can a research article be written and updated from such a workflow? Figures, tables and code are relatively easy to update, but what about descriptions of code, of methods, of data? What about a literature review, results and conclusions...can they also be updated?

We will provide details of our architecture, examples of its use and show a case study of a *live* research article that is created and maintained using these ideas.



## ABOUT THE AUTHOR(S)

(Both Gus Ellerm and Mark Gahegan are presenting)

**Mark Gahegan** is Professor in Computer Science at the University of Auckland, where he also directs the 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, 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 presented his work at the IEEE eScience'23 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.



# A Day in The Life (of eResearch)

Marcus Gustafsson  
The Centre for eResearch, University of Auckland  
[m.gustafsson@auckland.ac.nz](mailto:m.gustafsson@auckland.ac.nz)

## ABSTRACT / INTRODUCTION

This presentation looks at what 'BAU' means for a team of eResearch Specialists at a university. It will serve up a look at what might be expected from a modern service catalogue for eResearch in 2024 and then compare it with an everyday reality and what challenges the gaps pose. It will also touch on how an eResearch strategy might be created, what governance is required and how eResearch might be organised and supported within an institution.

## ABOUT THE AUTHOR(S)

- 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.



# Accelerating Geothermal Modelling with Cloud Computing

John O'Sullivan<sup>1</sup>, Ken Dekkers<sup>1</sup>, Michael Gravatt<sup>1</sup>, Mike O'Sullivan<sup>1</sup>, Joris Popineau<sup>1</sup>, Theo Renaud<sup>1</sup>,  
Jeremy Riffault<sup>1</sup>, and Ryan Tonkin<sup>1</sup>

<sup>1</sup>The University of Auckland, The Geothermal Institute, Auckland, New Zealand

<sup>2</sup>Amazon Web Services

[jp.osullivan@auckland.ac.nz](mailto:jp.osullivan@auckland.ac.nz)

Geothermal resources are expected to play an important role in helping to achieve a just, sustainable energy transition in New Zealand and around the world. Geothermal reservoir models are invaluable tools for managing the utilisation of geothermal resources. Traditionally these computationally intensive models are slow to run and difficult to manage. In 2019 the Geothermal Institute released a new, fast, parallelised geothermal simulator called Waiwera developed in partnership with GNS Science and funded by the Ministry of Business, Innovation and Employment. The simulator was developed with the support of New Zealand eScience Infrastructure (NeSI) and was benchmarked extensively on the Maui supercomputer. It has been shown to be orders of magnitude faster for solving typical geothermal models.

In 2022 we installed Waiwera on cloud computing infrastructure with support from Amazon Web Services (AWS). By creating virtual compute node clusters connected to AWS's fast FSx storage we have now been able to surpass Waiwera's performance measured on Maui for many of our standard models. The scalability of AWS's cloud computing platform means that running robust uncertainty quantification of geothermal models is now standard practice for many of our projects. Also, dedicated single compute nodes for smaller models, different types of nodes for pre- and post-processing models and remote desktop options available in different global regions are all significant benefits for migrating our workflow to the Cloud. The availability of local cloud computing resources in Indonesia also provides the perfect solution to data security constraints for our Indonesian clients.

NeSI's supercomputers will continue to provide invaluable support for our research and development. However, we expect that the flexibility, scalability, and power of AWS's cloud computing platform combined with next generation simulators like Waiwera will completely change the way the geothermal industry carries out modelling.



# Applying Large Language Models to Data Driven Research Platforms

Prof Richard Sinnott, Luca Morandini  
The University of Melbourne  
[luca.morandini@unimelb.edu.au](mailto:luca.morandini@unimelb.edu.au)

## ABSTRACT

Large language models (LLMs) such as chatGPT have the potential to massively augment the way in which data-driven research is conducted. This Melbourne eResearch Group at The University of Melbourne ([www.eresearch.unimelb.edu.au](http://www.eresearch.unimelb.edu.au)) is involved in numerous major data driven projects covering the complete gamut of research: from arts and humanities to zoology. This includes the Australian Research Data Commons supported:

- Australian Data Observatory ([www.ado.eresearch.unimelb.edu.au](http://www.ado.eresearch.unimelb.edu.au)), which has established the national social media aggregation platform for Australia, and
- the national bushfire data commons (<https://bdc poc.eresearch.unimelb.edu.au>), which incorporates extensive data sets related to bushfires in Australia and their impact on society.

This talk will present these platforms and importantly how LLMs offer capabilities that provide a step change in which data discovery and data analytics can be enhanced. The talk will cover the NeCTAR Cloud-based deployment of LLAMA2 – the open source LLM (<https://ai.meta.com/llama/>) from Meta (aka Facebook). The challenges and requirements of deploying LLMs to the Cloud are described, as well as the benefits of direct deployment of LLMs (rather than use of the openAI API) in terms of both the associated cost and importantly the privacy of end user data.

The talk will present a demonstration of the LLM solutions and how they support the two aforementioned platforms.

## ABOUT THE AUTHORS

**Professor Richard O. Sinnott** is Professor of Applied Computing Systems and Director of the Melbourne eResearch Group at the University of Melbourne. He has been lead software engineer/architect on an extensive portfolio of national and international projects, with specific focus on those research domains requiring finer-grained access control (security) and those dealing with big data challenges. He has over 450 peer reviewed publications across a range of applied computing research areas.

**Mr Luca Morandini** is a Cloud architect in the Melbourne eResearch Group. He was the data architect of the Australian Urban Research Infrastructure Network (AURIN) for 10+ years.



# Bridging the Digital Divide: Supporting Aotearoa's Journey to Digital Literacy Excellence

Dr Murray Cadzow, Dr Nisha Ghatak, Dr Tyler McInnes, and Dr Tom Saunders  
University of Otago, New Zealand eScience Infrastructure, Genomics Aotearoa, University of Auckland  
[murray.cadzow@otago.ac.nz](mailto:murray.cadzow@otago.ac.nz), [nisha.ghatak@nesi.org.nz](mailto:nisha.ghatak@nesi.org.nz), [tyler.mcinnnes@otago.ac.nz](mailto:tyler.mcinnnes@otago.ac.nz),  
[tom.saunders@auckland.ac.nz](mailto:tom.saunders@auckland.ac.nz)

## ABSTRACT / INTRODUCTION

Raising digital skills literacy in New Zealand is essential for the advancement and progress of the nation. As technology continues to transform every aspect of our lives, it is essential to ensure that researchers are equipped with the necessary skills to navigate the changing digital landscape effectively. Universities, research centres and crown research institutes across the motu strive to support a research workforce proficient in data science, high-performance computing, and advanced research methodologies, contributing significantly to the country's digital competitiveness and knowledge economy.

This Birds-of-a feather session aims to spotlight the diverse initiatives for digital upskilling led by organizations nationwide. It will encourage a focused dialogue on identifying the research requirements of the community as we prepare for next year's training programs. The session is geared towards addressing the training necessities of the research community, striving to meet their specific needs effectively.

## ABOUT THE AUTHOR(S)

- **Dr Murray Cadzow:** Murray is a Scientific Program mer within Research Teaching IT Support at the University of Otago. Prior to this he spent 11 years researching the genetic basis of gout and related diseases. Murray has been heavily involved in computational literacy and bioinformatic training at the University of Otago - organising Research Bazaar Dunedin and the Otago Bioinformatics Spring School. He is both a Carpentries instructor and instructor trainer. His teaching has focused on delivering digital literacy training to researchers, and the development and support of the local Carpentries community at Otago.
- **Dr Nisha Ghatak:** Nisha is the Research Communities Advisor and Training Lead at NeSI. She is also the Carpentries Regional Coordinator for Aotearoa New Zealand. and an Executive Council member for the Carpentries.
- **Dr Tyler McInnes:** Tyler is the Bioinformatics Training Coordinator for Genomics Aotearoa. He has research experience in the field of genetics, studying cancer epigenetics in colorectal cancer and limb and spinal development in *Xenopus laevis*. In his previous role as a Teaching Fellow, Tyler developed and implemented a series of workshops to support student learning, as well as new lecture content and practical labs with a focus on bioinformatics. He is now a certified Carpentries instructor
- **Dr Tom Saunders:** Tom has a background in entomology and biosecurity but is now an Engagement Specialist in the Centre for eResearch at the University of Auckland. Tom organises, hosts, and delivers digital research skills workshops, and helps researchers to manage their data throughout the research lifecycle. He collaborates on the organisation and delivery of ResBaz Aotearoa and cross-institution Carpentries events as a certified instructor.





# Bridging the Gap with OpenOnDemand

Jordi Blasco, Elisabeth Ortega-Carrasco, Christian Bustelo  
HPCNow!

[jordi.blasco@hpcnow.com](mailto:jordi.blasco@hpcnow.com), [elisabeth.ortega@hpcnow.com](mailto:elisabeth.ortega@hpcnow.com), [christian.bustelo@hpcnow.com](mailto:christian.bustelo@hpcnow.com)

## ABSTRACT / INTRODUCTION

As high-performance computing (HPC) resources become increasingly essential for scientific research, engineering, and computational-intensive tasks, there exists a growing need to democratize access to these powerful resources. This talk highlights the pivotal role of OpenOnDemand, a user-friendly web-based platform, in simplifying the transition from desktop computing to HPC clusters. OpenOnDemand offers an intuitive interface that enables desktop users, including those without extensive HPC expertise, to seamlessly access and leverage the capabilities of HPC clusters. In this talk, Jordi Blasco, CTO at HPCNow! will delve into the features and functionalities of OpenOnDemand, including the contributions made by the HPCNow! team in this vibrant OpenSource project. He will illustrate how this technology empowers scientists and engineers to harness the full potential of HPC resources for their computational workflows.

Open OnDemand incorporates a suite of cutting-edge technologies to facilitate seamless access and utilization of high-performance computing (HPC) resources. At its core, it leverages web-based interfaces, ensuring an intuitive and user-friendly experience. It harnesses Apache HTTP Server for robust web hosting and integrates with OAuth for secure authentication and authorization. Open OnDemand also employs technologies like Interactive Apps, Jupyter Notebooks, and RStudio for a flexible and collaborative computational environment.

Through a user-friendly environment, OpenOnDemand bridges the gap, making HPC accessible and maximizing the impact of computational research across multiple scientific domains. The case studies exposed in the talk emphasize the transformative impact of OpenOnDemand in democratizing HPC access and enabling a broader community to advance their research endeavors.

## ABOUT THE AUTHOR(S)

### - **Jordi Blasco Pallares**

As the Chief Technology Officer (CTO) with a background in computational physics, Jordi Blasco brings a unique perspective to HPCNow!, a consulting company focused on high-performance computing. With a passion for using technology to solve complex problems, Jordi has a proven track record of driving innovation and delivering results in diverse industries, including academia, government, and the private sector, from multiple Fortune 500 companies to a long tail of medium, small, and start-up institutions.

Jordi's expertise in computational physics has honed their skills in performance analysis, application tuning, modeling, and simulation, which they leverage to develop cutting-edge solutions for clients. He has a deep understanding of high-performance computing technologies, including parallel programming, performance analysis, application tuning, distributed computing, and cloud computing, and he works closely with their team to design and implement scalable, efficient, and cost-effective solutions.

As a leader, Jordi fosters a collaborative and innovative culture that encourages the team to think outside the box and push the boundaries of what's possible. He is committed to staying at the forefront of emerging technologies and trends in high-performance computing, and brings that knowledge and expertise to bear in his work with clients.

Whether he is helping a client optimize their workflow, accelerate their simulations, or develop new applications, Jordi's goal as CTO is to deliver solutions that exceed client expectations and drive real business value.

### - **Dr. Elisabeth Ortega Carrasco**

Dr. Elisabeth Ortega is the Research, Development, and Innovation Manager at HPCNow!. With her extensive STEM background and profound expertise in High-Performance



Computing (HPC), she is the driving force behind the human aspect of technological advancements. As a leader in R&D&I, her primary mission is to explore emerging trends in HPC utilization from both industry and academia. She endeavors to craft innovative solutions that streamline our customers' daily tasks, making them more efficient and cost-effective from a computational perspective.

Dr. Ortega's educational journey is multifaceted, encompassing a Ph.D. and an M.Sc. in Theoretical and Computational Chemistry, underpinned by her undergraduate degrees in chemistry and computer science. She is presently expanding her horizons by delving into quantum computing to propel HPC into the quantum era. This holistic knowledge base equips her with a unique perspective in the market, enabling her to foresee and capitalize on new opportunities, positioning us one step ahead in the industry.

Prior to her current role, Dr. Ortega dedicated seven years to designing and developing software for pharmaceutical companies. During this time, she not only created solutions but also educated customers on their usage. She also played a key role in sales, gaining valuable insights into how to effectively engage and serve users and customers, ultimately providing them with tailored solutions.

- **Christian Bustelo Yáñez**

Christian Bustelo is a professional with a solid background in computer science. He holds a degree in Computer Engineering with a specialisation in Computer Systems, which has provided him with knowledge of computer science and Computer Structures. In addition, he has a Master's degree in High Performance Computing (HPC), which further strengthens his expertise and interest in this field. In this Master's degree he has been nominated for an award for his job on his thesis, where he developed new functionalities related to Open OnDemand.

Since his early days in the world of computer science, Christian has shown an innate passion for this discipline. His interests lie in multiple areas, but since he started seeing specialised areas such as clusters and large-scale computing systems, he developed a genuine interest in them. His penchant for high-performance systems and his passion for innovation has made him a solid worker in his field.

Currently, Christian Bustelo is employed in the Research and Development department at HPC Now!, where he has been involved in a wide range of projects across various disciplines. His role primarily involves researching and developing solutions in the realm of high-performance computing. This position enables him to stay updated with the latest technologies and tools, as well as to explore and apply cutting-edge innovations in high-impact projects.



# Bringing Security in Tuakiri to Next Level

Vladimir Mencl  
Research and Education Advanced Network New Zealand  
[vladimir.mencl@reannz.co.nz](mailto:vladimir.mencl@reannz.co.nz)

## ABSTRACT / INTRODUCTION

For already over a decade, Tuakiri has been providing the NZ R&E community with a way to authenticate users in a trusted way with assurance about the user's attributes, as provided by the user's home organisation.

However, in the current state, services consuming a Tuakiri login only get the level of trust at a common baseline.

The security requirements are evolving and there is growing demand to bring the security up to a next level on several dimensions - even if applied only to a smaller selection of services and/or users.

Multi-factor authentication (MFA) has become the norm in many authentication scenarios. A number of organisations have enabled MFA also when authenticating users accessing their Tuakiri-connected Identity Provider (IdP). However, to deliver the value of MFA to a service using Tuakiri to authenticate users, we need a common way for SPs to request MFA and for IdPs to confirm MFA was used - and this should be independent of the specific MFA technology used. This need been addressed by the international R&E community with REFEDS MFA [1], and Tuakiri aims to adopt it. Some services need to see the strength of identity-proofing processes the user's identity has been verified to. The IdP can signal this by including relevant attribute values from the REFEDS Assurance Framework [2].

In case a security incident occurs, in the Federated Identity space, it is crucial to coordinate response among the multiple parties involved. The REFEDS SIRTFI framework [3] addresses this and provides a way for individual parties to signal they can follow the response processes outlined.

These enhancements would increase the value Tuakiri delivers to the community - and may also become required to access certain services.

This presentation will describe these security-posture improving changes in further detail and will outline what would be required of Tuakiri members to make use of these improvements.

### References:

[1] <https://refeds.org/profile/mfa>

[2] <https://refeds.org/assurance>

[3] <https://refeds.org/sirtfi>

## ABOUT THE AUTHOR(S)

Dr. Vladimir Mencl has been part of the New Zealand R&E community since 2006 and has been involved in identity and access management projects since the early days of the BeSTGRID project. When the Tuakiri project moved to REANNZ, Vlad joined REANNZ where he is part of the Systems team as a Lead Software Engineer.



# Building custom web portals for HPC and mid-tier computing

Blair Bethwaite, Thomas Berger  
New Zealand eScience Infrastructure – NeSI, Auckland, AKL, New Zealand  
[blair.bethwaite@nesi.org.nz](mailto:blair.bethwaite@nesi.org.nz), [thomas.berger@nesi.org.nz](mailto:thomas.berger@nesi.org.nz)

## ABSTRACT / INTRODUCTION

### Background

High-Performance Computing (HPC) and mid-tier computing play crucial roles in scientific research. To overcome the challenges associated with accessing and managing these resources, customised web portals have emerged as a viable solution. This abstract presents our approach to developing custom web portals for HPC and mid-tier computing environments, leveraging the power of Open OnDemand (OOD).

### Actions

Our development process involved utilising OOD, an open-source platform that simplifies access to HPC resources through web-based portals. We engaged in comprehensive user requirement analysis, collaborating with domain experts and end-users to identify their specific needs. Building upon OOD's capabilities, we designed intuitive user interfaces with features such as job submission, data management, and result visualisation. By integrating OOD with sophisticated backend systems, we ensured efficient management and scalability of complex computational tasks. Agile methodologies enabled us to rapidly prototype, gather user feedback, and iteratively enhance the web portals.

### Results

The implementation of customised web portals powered by OOD has yielded significant onboarding improvements for both users and support staff. Researchers and PhD students can seamlessly interact with HPC and mid-tier computing infrastructures, benefiting from the simplified access provided by OOD.

### Conclusion

Customised web portals have empowered researchers and PhD students to harness the potential of HPC and mid-tier computing resources. The iterative development process, coupled with OOD's user-centric design principles, has ensured close alignment with end-users' needs.

## ABOUT THE AUTHOR(S)

- Blair Bethwaite

Blair is the NeSI Solutions Manager responsible for user support services on the NeSI platforms, from onboarding and application support to optimisation of the science codes that are run on the Platforms.

- Thomas Berger

Thomas 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.



# Co-designing machine-actionable Data Management Planning (maDMP) using ReDBox at the University of Auckland.

Yvette Wharton<sup>1</sup>, Simon Esling<sup>1</sup>  
1Waipapa Taumata Rau | University of Auckland  
[y.wharton@auckland.ac.nz](mailto:y.wharton@auckland.ac.nz); [s.esling@auckland.ac.nz](mailto:s.esling@auckland.ac.nz)

## ABSTRACT / INTRODUCTION

Research data management is an important part of any research project. It helps the researcher/research team collaborate on the data lifecycle, from collection to storage to preservation and everything in between. It also outlines a structured approach to working with data and aligns the research project with infrastructure needs, resources, and valuable services. But how readily and easily does a methodical approach to data management occur in complex research environments? There is a tacit understanding of the benefits of a data management plan, but in practice, data management planning can too often be considered a burdensome administrative process. This presentation will discuss the co-design approach we have utilised at Waipapa Taumata Rau, from gathering user requirements, narrowing down and selecting ReDBox as the Data Management Planning tool to pilot and designing resources and templates. Alongside user preferences, a particular focus was the ability to provide and cascade information to/from other research-related systems such as:

- Researchers can request services, such as storage or virtual computing, initiate an ethics application or produce a metadata record for institutional Figshare.
- Researchers can create Data Artefact Records from their DMPs, and, from these records, they can produce findable Data Publication Records.
- For early career researchers the guidance and resources also serve as a pathfinder for valuable research-related and data management information as each section of the Data Management Plan is completed.

The maDMP template layout and related guidance continues to be an iterative, co-design process to ensure the platform is relevant for various researcher requirements. The long-term intention is to continue to create connections between different platforms, systems, and services within the University to further reduce administrative burden whilst increasing research data management maturity.

## ABOUT THE AUTHOR(S)

Yvette Wharton is the eResearch Solutions Lead at the Centre for eResearch, University of Auckland, working on research data management services and researcher enablement projects.

<http://orcid.org/0000-0002-6689-8840>

Simon Esling is a Senior eResearch Engagement Specialist at the Centre for eResearch, University of Auckland, working on co-developing a machine action Data Management Plan in the context of a Research Data Management Programme.

<https://orcid.org/0000-0001-5994-7619>



# Collaboration Cafe: Sharing sensitive research data infrastructure architectures BoF

Martin Feller<sup>1</sup>, Andre Geldenhuis<sup>1</sup>, Yvette Wharton<sup>1</sup>, Bincy Jacob<sup>1</sup>  
<sup>1</sup>Waipapa Taumata Rau | University of Auckland, Tāmaki Makaurau Auckland,  
Aotearoa New Zealand

[m.feller@auckland.ac.nz](mailto:m.feller@auckland.ac.nz), [andre.geldenhuis@auckland.ac.nz](mailto:andre.geldenhuis@auckland.ac.nz), [y.wharton@auckland.ac.nz](mailto:y.wharton@auckland.ac.nz), [bincoy.jacob@auckland.ac.nz](mailto:bincoy.jacob@auckland.ac.nz)

## ABSTRACT / INTRODUCTION

Across universities and institutes, there is a diversity of data types and workflows, including research involving sensitive data. These data are commonly subject to additional legal, ethical and other obligations which impose restrictions on access, use, and handling. However, across NZ/Aotearoa research institutions, we currently have few shared or adopted policies, architectures or practices for these data, leading to difficulty for researchers and their institutions aiming to support them.

We will present a summary of international best practices for sensitive research data environments alongside a few institutional use cases. We invite participants to share their experiences to progress a combined vision for sensitive data research infrastructure in Aotearoa.

Together, we will explore possible issues and opportunities for alignment and collaboration of architectural patterns and best practices, such as:

- cybersecurity considerations
- roles and responsibilities
- identity and access management
- application components
- data accessibility

This session is aimed at those providing research infrastructure services and support across sensitive data management (e.g., eResearch, research management, research IT, digital research skills, and policy) across the research sector.

## ABOUT THE AUTHOR(S)

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.

**Andre Geldenhuis** works as an eResearch Solutions Specialist at the Centre for eResearch, University of Auckland, Andre is currently working on the Secure Research Environment initiative developing and supporting pilots using AWS-TREEHOSE and a local Nectar implementation.

**Yvette Wharton** is the eResearch Solutions Lead at the Centre for eResearch, University of Auckland, working on research data management services and researcher enablement projects.

<http://orcid.org/0000-0002-6689-8840>

**Dr Bincy Jacob** is the Snr eResearch Engagement Specialist working for the RDM programme at Waipapa Taumata Rau. Coming from a background of mass spectrometry and proteomics, she is now working with different communities within university to develop a secure research environment (SRE) for the management of sensitive data.



# Collaboration Cafe: Sharing sensitive research data management practices (BoF)

Yvette Wharton<sup>1</sup>, Laura Armstrong<sup>1</sup>, Bincy Jacob<sup>1</sup>, Sarah Hopkins<sup>1</sup>, Claire Rye<sup>2</sup>  
<sup>1</sup>Waipapa Taumata Rau | University of Auckland, Tāmaki Makaurau Auckland,  
Aotearoa New Zealand

<sup>2</sup>New Zealand eScience Infrastructure, Auckland, Auckland, New Zealand

[y.wharton@auckland.ac.nz](mailto:y.wharton@auckland.ac.nz), [l.armstrong@auckland.ac.nz](mailto:l.armstrong@auckland.ac.nz), [binco.jacobs@auckland.ac.nz](mailto:binco.jacobs@auckland.ac.nz), [s.hopkins@auckland.ac.nz](mailto:s.hopkins@auckland.ac.nz),  
[claire.rye@auckland.ac.nz](mailto:claire.rye@auckland.ac.nz)

## ABSTRACT / INTRODUCTION

Across universities and institutes, there is a diversity of data types and workflows, including research involving sensitive data. These data are commonly subject to additional legal, ethical and other obligations which impose restrictions on access, use, and handling. However, across NZ/Aotearoa research institutions, we currently have few shared or adopted policies, architectures or practices for these data, leading to difficulty for researchers and their institutions aiming to support them.

We will present a summary of international developments for sensitive research data management and a few institutional use cases. Participants will be invited to also share their experiences to progress towards a shared vision of the policy, practice and training for sensitive research data. Together, we will explore issues and opportunities for alignment and collaboration, including:

- national and institutional RDM policies
- data classification
- sensitive data sharing
- training needs
- researcher support

This session is aimed at those providing research services and support across sensitive data management (e.g., eResearch, research management, research IT, digital research skills, and policy) across the research sector.

## ABOUT THE AUTHOR(S)

Yvette Wharton is the eResearch Solutions Lead at the Centre for eResearch, University of Auckland, working on research data management services and researcher enablement projects.  
<http://orcid.org/0000-0002-6689-8840>

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. <http://orcid.org/0000-0003-2370-3924>

Dr Sarah Hopkins is an eResearch Engagement Specialist in the Centre for eResearch, Waipapa Taumata Rau | University of Auckland. She has a clinical research background in exercise physiology and youth mental health but is now working to support researchers to navigate the evolving requirements for robust research data management. <https://orcid.org/0000-0002-4705-5362>

Dr Bincy Jacob is the Snr eResearch Engagement Specialist working for the RDM programme at Waipapa Taumata Rau. Coming from a background of mass spectrometry and proteomics, she is now working with different communities within university to develop a secure research environment (SRE) for the management of sensitive data.

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. Most recently, she was a Product Owner for the Ingestion service of the Human Cell Atlas Data Coordination Platform, overseeing the development of the software infrastructure and metadata standards that supports data sharing across HCA globally, based at the European Bioinformatics Institute (EMBL-EBI). <https://orcid.org/0000-0003-4630-7836>





# Connecting Researchers Anywhere

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

## ABSTRACT / INTRODUCTION

The underlying purpose of REANNZ is to provide the National Research and Education Network (NREN) for Aotearoa. NRENs are specialised internet providers established in over 120 countries to help researchers and educators study, research and collaborate. Historically, this has been focussed on enabling the transfer of large data sets over fibre optic networks at high speeds.

REANNZ sought feedback from our members about areas in which they needed more capability and one of the areas was to help them deliver a “consistent digital experience” wherever users are located, within Aotearoa and beyond.

The telecommunications industry has a number of emerging communications technologies that can be leveraged by researchers to access digital services, or transfer data from remote locations that have traditionally not been able to be connected to the REANNZ network.

A proof of concept has been completed using 4G mobile broadband, but the capability has been designed to make use of any telecommunications technology that can provide a connection to the Internet. This includes satellite technologies, 5G mobile broadband and Internet of Things (IOT) focussed technologies such as LoRaWAN.

Enabling secure digital communications from remote locations provides a tool for researchers to leverage for data collection, or access digital services in the field.

## ABOUT THE AUTHOR(S)

- David Brownlie
- David 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.



# Creating new futures with data

jan.sheppard@esr.cri.nz

## ABSTRACT / INTRODUCTION

We are operating in a time of exponential change, largely driven by technology. This gives us the opportunity to use data to experiment into the future to see beyond the horizon so we can make the best decisions today for tomorrow. Jan will explore what this looks like in reality, and draw on examples from her work at the Institute of Environmental Science and Research where such technologies combined with data are being applied to address the big issues of the future for Aotearoa, including climate change resilience and pandemics.

## ABOUT THE AUTHOR(S)

Name – Jan Sheppard

Job title – Chief Data & Analytics Officer

Organisation – Institute of Environmental Science and Research

Bio –

Jan Sheppard is an innovator who uses data to build new worlds, to see things not previously visible and to enable organisations to be better versions of themselves. Jan has extensive experience across a range of sectors, taking a system perspective to use data to drive transformational change. Jan is the Chief Data & Analytics Officer at ESR, where she heads up the data science, computational science, business intelligence and information management practices.

Jan's approach and the outcomes achieved were acknowledged globally in 2020 and again in 2022, when she was named among the top 100 innovators in data and analytics in the world.



# DeepWeather: using deep learning to produce high-resolution weather forecasts over New Zealand

Emily O’Riordan<sup>1</sup>, Greg Bodeker<sup>1</sup>, Leroy Bird<sup>1</sup>, Greg Pearson<sup>2</sup>, Mark Schwarz<sup>2</sup>, Chris Noble<sup>2</sup>, Tui Warmenhoven<sup>3</sup>  
1. Bodeker Scientific, 2. MetService 3. He Oranga mō ngā Uri Tuku Iho Trust  
[emily@bodekerscientific.com](mailto:emily@bodekerscientific.com)

## ABSTRACT / INTRODUCTION

New Zealand is vulnerable to extreme weather events due to its complex climate and terrain. The ability to forecast the weather in more detail at local scales would better predict many of these events, allowing for enhanced warnings and preparations. However, producing these high-resolution forecasts using physics-driven numerical weather prediction (NWP) models is prohibitively expensive. The DeepWeather project aims to produce a generative deep learning (DL) model, taking low-resolution NWP forecasts as inputs, and outputting an ensemble of high-resolution forecasts. To train the model, we leverage domestic high-performance computing resources thanks to NeSI, allowing for faster training, as well as experimentation between multiple DL architectures and parameterisations. The DL model will be inexpensive to run once it has been trained, allowing us to produce high-resolution forecasts at a fraction of current costs.

The DeepWeather project team are working with hapū in Tairāwhiti to gain an understanding of how localised weather forecasts can be designed to best cater to those affected by extreme weather events. A rain-monitoring project will be created in collaboration with ākonga in the area to encourage interest in weather science and provide localised data for the training of our DL model. This project is in collaboration with MetService and, once trained, the DL model will be embedded into MetService’s NWP model through two-way coupling. This will allow the NWP model outputs to be instantly fed into the DL model, generating high-resolution forecasts at minimal cost, which can then be used to nudge low-resolution NWP simulations. As the DL model will resolve sub-grid-scale processes and account for local terrain, the DeepWeather model (that fuses the NWP model outputs and the DL model) should enhance the accuracy of weather forecasts over New Zealand.

## ABOUT THE AUTHOR(S)

Dr Emily O’Riordan

Emily is a postdoctoral researcher with Bodeker Scientific, a climate and weather research organisation based in Otago. She is investigating how artificial intelligence may be used to improve both the accuracy and cost of producing high-resolution forecasts over Aotearoa New Zealand. Her research encompasses neural network architecture, numerical weather prediction and the impact of more detailed weather forecasts on vulnerable communities. This project works closely with meteorologists at MetService and other climate scientists both domestically and internationally. She recently received her PhD from Cardiff University’s School of Mathematics, where she studied measures of proximity and decorrelation for high dimensional data. Her PhD project produced several new data science methods, suitable for applications across a variety of disciplines.



# Do I Need a DOI?

Andrea Goethals  
National Library of New Zealand  
[Andrea.Goethals@dia.govt.nz](mailto:Andrea.Goethals@dia.govt.nz)

## ABSTRACT / INTRODUCTION

This lightning talk will be a short introduction to DOIs (Digital Object Identifiers). John Kunze (2021) [1] describes ten persistent myths about persistent identifiers. Similarly, in this lightning talk I will describe some of the misunderstandings I have encountered in my discussions with people about DOIs in my role coordinating the NZ DOI Consortium.

References:

[1] John Kunze 2021, ARK Alliance website, ARK Alliance Key (ARK), accessed 27 October 2023, <https://arks.org/blog/ten-persistent-myths-about-persistent-identifiers/>.

## ABOUT THE AUTHOR(S)

Andrea Goethals started her digital preservation career in 2003 as a computer scientist working on the technical challenges. Since then she has worked in many different roles focusing on the policies, strategies and people that make digital preservation programmes possible. She is now Manager of Digital Preservation and Data Capability at the National Library of New Zealand. She participates in many local, regional and international working groups including NZ DOI Consortium, Australasia Preserves, IIPC Steering Committee, iPres Steering Group, Digital Preservation Storage Criteria WG, DataCite CESC, NSLA DPN and DPC-Australasia.



# Harmony in Progress: Research Lifecycle Maturity Model & Digital Transformation at the University of Otago

Rasha Abu Safieh  
University of Otago  
[rashaabusafieh.alfar@otago.ac.nz](mailto:rashaabusafieh.alfar@otago.ac.nz)

## ABSTRACT / INTRODUCTION

Embark on a transformative journey towards a future of unparalleled Research Excellence at the University of Otago. This enlightening session introduces the Stakeholder-Engagement-Centric approach, meticulously applied to forge Otago University's inaugural Research Lifecycle Maturity Model. Explore the intricacies of this groundbreaking methodology, which not only sharpens research capabilities but also propels the University into a future defined by digital excellence. Join us in highlighting Otago's steadfast dedication to digitally empowering its research community, in perfect harmony with the strategic vision outlined in its Pae Tata Strategic Plan 2030..

## ABOUT THE AUTHOR(S)

In her capacity as the Acting Operations Strategist and Digital Projects Manager at Otago University, Rasha Abu Safieh spearheads transformative initiatives that redefine digital paradigms and operational models. Recognized as a Time Magazine Next Generation Leader in 2018, she earned this distinction for her decade-long dedication to social entrepreneurship within one of the most challenging political and economic contexts. Rasha has successfully crafted innovative business models, bringing a wealth of experience in complex systems leadership, systems mapping, and inclusive entrepreneurship to drive impactful change within the institution.

As the Founder in Residence of STARlab (Systems Transformation Action Research Lab), Rasha applies her unique expertise in systems design and management to oversee and revolutionize research projects. A devoted Edmund Hillary Fellow, she is unwavering in her commitment to shaping a better world for the next generation, particularly in Aotearoa, New Zealand. In her role as the Co-Chair of the Catalyst2030 New Zealand Chapter, Rasha fervently advocates for social justice, leading a community of change-makers dedicated to achieving Sustainable Development Goals. Her multifaceted expertise, coupled with her pivotal position at Otago University and her role as Founder in Residence of STARlab, seamlessly aligns with the institution's dedication to Research Excellence and Digital Innovation. Rasha Abu Safieh stands as a dynamic force, propelling positive change and embodying a commitment to advancing societal well-being.



# Fishing for parallelisation strategies

Alexander Pletzer<sup>1</sup>, Inna Senina<sup>2</sup> and Chris Scott<sup>3</sup>  
<sup>1</sup>NeSI/NIWA, <sup>2</sup>The Pacific Community, <sup>3</sup>NeSI/University of Auckland  
[alexander.pletzer@nesi.org.nz](mailto:alexander.pletzer@nesi.org.nz)

## ABSTRACT / INTRODUCTION

Fishing is a major economic source for many Pacific Islands. SEAPODYM is a modelling framework designed to predict tuna population dynamics under fishing pressure and various environmental conditions (e.g., climate change). The code integrates spatiotemporal, age-structured advection-reaction-diffusion models with statistically fitted observation data. Simulations can take on the order of 10 hours, depending on the grid resolution and the number of fish cohorts. To reduce the wall clock time, NeSI was approached by the Pacific Community to review the code and make suggestions for applying parallelisation and other acceleration strategies. Here, we present our findings, including the development of an MPI toy program that emulates the behaviour of SEAPODYM to obtain an estimate for the maximum achievable parallel scalability of the code on NeSI's parallel platform.

## ABOUT THE AUTHOR(S)

- Alex Pletzer
- Alex is a research software engineer who helps researchers run better and faster on NeSI platforms. On his spare time, Alex windsurfs, paints and plays ping pong.



# Flooding an HPC: parallelism optimisation in the STRAND project

Maxime Rio<sup>1,2</sup>, Murray Cadzow<sup>3</sup>, Quyen Nguyen<sup>3</sup>  
<sup>1</sup>NeSI, <sup>2</sup>NIWA, <sup>3</sup>University of Otago

[maxime.rio@nesi.org.nz](mailto:maxime.rio@nesi.org.nz), [murray.cadzow@otago.ac.nz](mailto:murray.cadzow@otago.ac.nz), [quyen.nguyen@otago.ac.nz](mailto:quyen.nguyen@otago.ac.nz)

## ABSTRACT / INTRODUCTION

The STRAND Marsden Fund Project is an interdisciplinary project exploring climate-change flooding related risks to residential property values across space and time, and the related implications for financial stability.

One workflow of the STRAND project aims at evaluating the flooding risks for Dunedin properties under various climate change scenarios, architecture choices and the interaction with data uncertainty. This workflow involves generating and aggregating thousands of simulations across many properties. The embarrassingly parallel nature of the problem made it a good fit for a high-performance computing (HPC) platform.

In this talk, we will explore how RTIS (University of Otago) and NeSI partnered to optimise this code. After emphasizing the numerical challenges, we will highlight why and how we combined a workflow management system, Snakemake, and a parallel toolbox, Dask, to distribute tasks efficiently and limit bottlenecks, while keeping the solution portable across two different HPC platforms.

## ABOUT THE AUTHOR(S)

Dr Maxime Rio is a data science engineer and data scientist at NeSI and NIWA. He enjoys helping researchers to analyse their data, from visualisation to machine learning and probabilistic modelling.

Dr Murray Cadzow is a Scientific Programmer within Research Teaching IT Support (RTIS) at the University of Otago. Prior to this he spent 11 years researching the genetic basis of gout and related diseases. Murray has been heavily involved in computational literacy and bioinformatic training at the University of Otago - organising Research Bazaar Dunedin and the Otago Bioinformatics Spring School. He is both a Carpentries instructor and instructor trainer. His teaching has focused on delivering digital literacy training to researchers, and the development and support of the local Carpentries community at Otago.

Dr Quyen Nguyen is the STRAND Marsden Fund Project Postdoctoral Fellow (2021–2024) hosted at the School of Surveying, University of Otago. She is the modeller for the STRAND Marsden Fund Project entitled "Should I stay or should I go? Climate-change risks to property values across space and time, and the related implications for financial stability". Dr Nguyen is also working with GNS Science as a Climate Change Economist. Her research interests are in climate finance and data science.



# Handling Health Data: FAIR Research Objects for Trusted Research Environments

Carole Goble  
The University of Manchester, UK, ELIXIR-UK, Health Data Research UK  
[carole.goble@manchester.ac.uk](mailto:carole.goble@manchester.ac.uk)

## ABSTRACT / INTRODUCTION

Trusted Research Environments (TREs) are secure locations in which data are placed for researchers to analyse. TREs can be set up to host administrative data, hospital data or any other data that needs to remain securely isolated. It is hard for a researcher to perform an analysis across multiple TREs, requesting and gathering the data needed from each one. Federated analysis widens the scope of research and makes more effective use of data, but that data needs to be analysed across geographical or governance boundaries, for example in devolved healthcare in the UK and across national borders in Europe.

A federated infrastructure makes it much easier for analysis tools to access multiple TREs. Health Data Research UK (<https://www.hdruk.ac.uk/>) through its DARE UK programme (<https://dareuk.org.uk/>) is developing a blueprint for TRE federation [1] and tools for federated data discovery. ELIXIR, the European Research Infrastructure for Life Science Data (<https://elixir-europe.org/>), has developed Federated European Genome-Phenome Archive (FEGA) [2] and services for FAIR data management and computational workflows using GA4GH standards [3].

There are different ways of implementing the well-established TREs, and many popular analysis tools already in widespread use, so solutions need to be readily adoptable by existing systems. Moreover, the infrastructure needs to work within the “Five Safes” framework [4] that aims to protect data and enable data services to provide safe research access to data. The “Five Safes RO-Crate” [5] is a new way of packaging up the digital objects needed for research requests and results with the information needed for the tools and TRE providers to ensure that the Crates are reviewed and processed according to Five Safes principles. RO-Crate [6] is a community effort to establish a lightweight, native approach to packaging research data with their metadata (<https://www.researchobject.org/ro-crate/>). Sponsored by ELIXIR and others, it has become a widely adopted framework for inter-service exchange, resource archiving, and reproducible reporting, used by digital research infrastructures and their services, including ELIXIR, the European Open Science Cloud, and the Australian BioCommons. It is an implementation of the FDO Forum’s FAIR Digital Objects (<https://fairdo.org/>). The TRE-FX project (<https://trefx.uk/>) has piloted FAIR Five Safes RO-Crates and answering data queries within HDR UK TREs using pre-approved workflows using ELIXIR’s workflow execution technologies. Partnering with TREs from Scotland, Wales and England and analysis toolkits (DataSHIELD, BitFount), TRE-FX streamlines the exchange of requests and results between analysis clients and TREs while ensuring that the access is safe and the process transparent. TELEPORT (<https://dareuk.org.uk/driver-project-teleport/>), a sister DARE UK project, follows a complementary federation strategy of ethereal “pop-up” TREs for requests that are only feasible over combined TREs. The combination of TRE-FX and TELEPORT is a powerful hybrid capable of addressing practical federated analysis patterns working within current data governance processes.

From March 2024 HDR-UK and ELIXIR will combine forces in the Horizon Europe EOSC-ENTRUST project which aims to create a European network of Trusted Research Environments for sensitive data and to drive European interoperability by joint development of a common blueprint for federated data access and analysis.

## References

[1] DARE UK, “Federated Architecture Blueprint”, 2023, <https://dareuk.org.uk/our-work/federated-architecture-blueprint/>





- [2] European Genome-Phenome Archive, FederatedEGA, <https://ega-archive.org/federated>
- [3] Thorogood A et al, “International federation of genomic medicine databases using GA4GH standards”, *Cell Genomics*, 1(2), 2021, <https://doi.org/10.1016/j.xgen.2021.100032>
- [4] UK Office of National Statistics, The ‘Five Safes’ – Data Privacy at ONS, 2017, <https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/>
- [5] Soiland-Reyes, S., Wheeler, S., Giles, T., Goble, C., & Quinlan, P. (2023). TRE-FX Technical Documentation - Five Safes RO-Crate (0.4). Zenodo. <https://doi.org/10.5281/zenodo.10376350>
- [6] Soiland-Reyes S, Sefton P, Crosas M, Castro LJ, Coppens F, Fernández JM, Garijo D, Grüning B, Rosa ML, Leo S, Ó Carragáin E, Portier M, Trisovic A, RO-Crate Community, Groth P, Goble C (2022):  
“Packaging research artefacts with RO-Crate” *Data Science* 5(2), <https://doi.org/10.3233/DS-210053>

## **ABOUT THE AUTHOR(S)**

Carole Goble CBE FREng FBCS

Carole Goble is a Professor of Computer Science at the University of Manchester, UK. She is a leader in Digital Research Infrastructures, translating technical innovations in distributed computing, semantic and metadata technologies, data and software sharing and computational workflows into FAIR and Open information solutions for scientists, in particular the Life Sciences and Biodiversity. She is currently: Joint Head of Node of ELIXIR-UK the UK node of ELIXIR, the European Research Infrastructure for Life Science Data; joint lead of the Federated Analytics programme for Health Data Research UK and a founder of the UK’s Software Sustainability Institute. Carole is an author of the seminal FAIR principles for scientific data and recipient of the Microsoft Jim Gray award for her contributions to eScience.



# Hard Questions for Soft Skills: Carpentries Community Collaboration in Aotearoa & Australia

Liz Stokes, Australian Research Data Commons, [liz.stokes@ardc.edu.au](mailto:liz.stokes@ardc.edu.au),  
Dr Nisha Ghatak, New Zealand eScience Infrastructure

## ABSTRACT / INTRODUCTION

### Situation

In July 2023 we celebrated the 25th Anniversary of the first Software Carpentry workshop. Researchers from across The Carpentries community reflected on how their scientific practices and trainer expertise have benefited from engaging with peers from around the world. What started out as “gapfiller” for technical skill development has evolved to represent best practice in open research methods, pedagogical practice and community development. (OECD, 2020) But how do we evaluate our community efforts?

A skilled research workforce relies on our ability to build capacity in data and computational skills for conducting efficient, open, and reproducible research. Capacity is strengthened by the extent to which soft skills are applied in community development (Williams et al, 2023). The Carpentries strategy for teaching openly licenced curricula grounded in evidence-based train-the-trainer program and community development has many benefits for both Aotearoa (New Zealand) and Australian research communities. Importantly, our countries are close and different enough to make comparison and collaboration fruitful endeavours. Initiated in 2023 by NeSI and ARDC, the AU-NZ Carpentries community calls have brought to the surface many shared challenges, providing opportunities for skill sharing and problem solving.

### Task

Successful community development requires candid evaluation of setbacks and opportunities, and an open dialogue with community members. Foregrounded by feedback from our researcher communities, this BoF will review community development efforts and strategic partnerships across the eResearch sector undertaken in both countries.

### Action

This Birds of a Feather session invites delegates to share experience with community evaluation. An adapted engagement model will follow, reflecting on evaluative practices for community development. BoF participants will discuss which skill development and community strategies resonate. Recommendations for new opportunities and target communities will be sought.

### Result

Synthesis of the BoF discussions and recommendations will inform strategic development and future efforts to nourish the Trans-Tasman Carpentries community. We seek to strengthen our networks by asking the hard questions of how and why we foster community.

OECD (2020), "Building digital workforce capacity and skills for data-intensive science", OECD Science, Technology and Industry Policy Papers, No. 90, OECD Publishing, Paris, <https://doi.org/10.1787/e08aa3bb-en>

Williams, J. J., Tractenberg, R. E., Batut, B., Becker, E. A., Brown, A. M., Burke, M. L., ... & Woodley, L. (2023). Optimizing Short-format Training: an International Consensus on Effective, Inclusive, and Career-spanning Professional Development in the Life Sciences and Beyond. *bioRxiv*, 2023-03.

## ABOUT THE AUTHOR(S)

**Liz Stokes** is a Skills Development Lead in Trainer Communities for the Skilled Workforce & Development Team at the Australian Research Data Commons. Liz is a keen advocate for social infrastructure and facilitating learner communities. They coordinate the ARDC Carpentries Partnership and are active in the global Carpentries community. <https://orcid.org/0000-0002-2973-5647>

**Dr Nisha Ghatak** is a Research Communities Advisor - Training Lead at NeSI. 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 also serves as a community elected Executive Council member for The Carpentries.





# Indigenous Data Sovereignty Checks and Permissions for Code Repositories

David Eyers  
University of Otago  
[David.Eyers@otago.ac.nz](mailto:David.Eyers@otago.ac.nz)

## ABSTRACT / INTRODUCTION

During software development work within the Science for Technological Innovation (SfTI) Veracity Spearhead, we noted that Git repositories that we were using needed higher quality metadata to support indigenous data sovereignty.

We are experimenting with the addition of another metadata file, aiming to have it join the existing set of such files (e.g., including README, LICENSE, COPYING, etc.).

The content of this new metadata file would be a human-readable description of the provenance of the repository's contents, with a specific emphasis on the community and stakeholder discussions and agreements undertaken. We believe that this notion of provenance and description of acceptable use is orthogonal to the existing notions of licensing, and deserves specific highlight (as opposed to burying such details within the project README).

Despite the contents of the file being aimed at human audiences, we are also aiming to support content that can be understood by human readers, but also be automatically parsed by software.

Examples include:

- Making reference to Traditional Knowledge labels (TK-labels; introduced and managed by the Local Contexts organisation).
- Referencing permission (or not) to treat the repository as a bit-stream, and replicate it into cloud storage outside of Aotearoa New Zealand.

Note that the provenance of this abstract submission is not yet complete with the likely addition of further authors and citations.

## ABOUT THE AUTHOR(S)

David Eyers is a Professor in the School of Computing at the University of Otago. He has broad research interests including secure and appropriate use of cloud computing, and the tracking of data provenance.



# Large Language Models in Scientific Research

Ian Foster  
[foster@anl.gov](mailto:foster@anl.gov)

## ABSTRACT / INTRODUCTION

### Large Language Models in Scientific Research

This talk delves into the role of Large Language Models (LLMs) like GPT-4 in scientific research. I explore how such models can revolutionize scientific methodologies by processing and synthesizing extensive scientific literature, aiding in hypothesis generation and data analysis. I examine practical applications of LLMs in science, showcasing their use in experimental design, and how they may contribute to democratization of scientific knowledge by making complex information more accessible. I also review challenges like accuracy, bias, and ethical concerns in scientific knowledge generation and dissemination, and propose strategies like model fine-tuning and validation protocols to mitigate these issues. I conclude by considering how we may use LLMs to build intelligent agents that work in partnership with human scientists to accelerate discovery.

Ian Foster is Senior Scientist and Distinguished Fellow, and director of the Data Science and Learning Division, at Argonne National Laboratory, and the Arthur Holly Compton Distinguished Service Professor of Computer Science at the University of Chicago. He has a BSc degree from the University of Canterbury, New Zealand, and a PhD from Imperial College, United Kingdom, both in computer science. His research is in distributed, parallel, and data-intensive computing technologies, and their applications to scientific problems. He is a fellow of the AAAS, ACM, BCS, and IEEE, and has received the BCS Lovelace Medal; IEEE Babbage, Goode, and Kanai awards; and ACM/IEEE Ken Kennedy award.

## ABOUT THE AUTHOR(S)

Ian Foster is Senior Scientist and Distinguished Fellow, and director of the Data Science and Learning Division, at Argonne National Laboratory, and the Arthur Holly Compton Distinguished Service Professor of Computer Science at the University of Chicago. He has a BSc degree from the University of Canterbury, New Zealand, and a PhD from Imperial College, United Kingdom, both in computer science. His research is in distributed, parallel, and data-intensive computing technologies, and their applications to scientific problems. He is a fellow of the AAAS, ACM, BCS, and IEEE, and has received the BCS Lovelace Medal; IEEE Babbage, Goode, and Kanai awards; and ACM/IEEE Ken Kennedy award.



# Leveraging our greatest unique resource, University of Otago student internship program

Wallace Chase  
University of Otago  
[wallace.chase@otago.ac.nz](mailto:wallace.chase@otago.ac.nz)

## **ABSTRACT / INTRODUCTION**

When it comes to staffing and talent what is the one unique resource that universities have but is often underutilized? Learn about how the University of Otago leverages student employment and internships to grow the next generation of IT professionals. For the last 3 years University of Otago has given opportunities to students to work in roles from the support desk to network engineering to IT Security. With plans to greatly expand on the successful program learn what has made it a valuable success for both the students and the full time university staff.

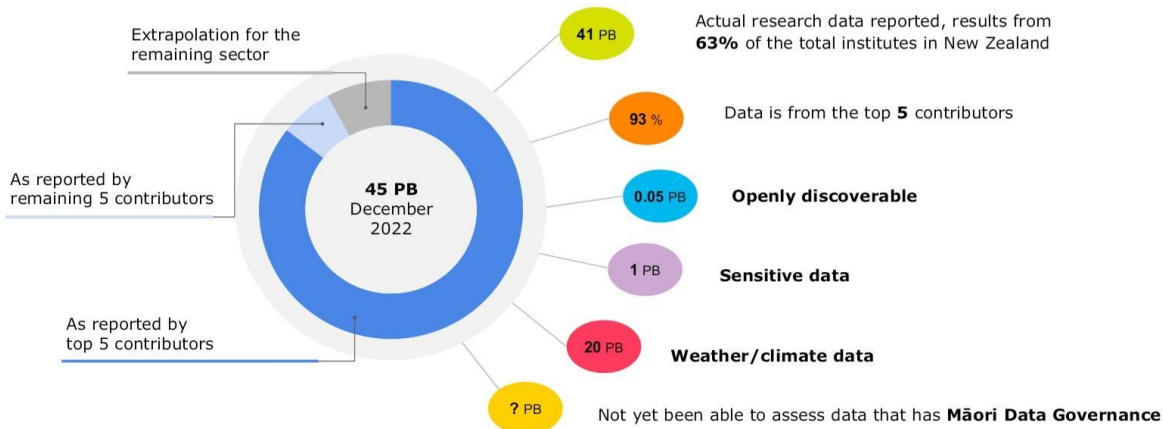


# Making research data count

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

## ABSTRACT / INTRODUCTION

Research Data is considered a valuable asset by Researchers and their Institutions alike. Yet, as institutions and infrastructures we do not report on data volume and other characteristics about our data holdings which begs further consideration about its value and the culture we have around managing research data. Working with Research Data Culture Conversation (RDCC) colleagues from Australia, in preparation for eResearch NZ 2023 we started asking a series of apparently simple questions to the Aotearoa New Zealand research data community. This work has produced the first 'Macro View' of Research Data as at December 2022. Giving us both a number, 45 PB (Petabytes) - an estimate of the total research data holdings in Aotearoa New Zealand, and a series of observations, such as how difficult this exercise was or of how the measurement and valuing of research data is mostly absent. We report that less than 0.1 % of our data volume is openly discoverable.



This BoF aims to bring the community of data holders together. Specifically, we anticipate working with the community to define what it is about these data that we might want to know and report on together.

Questions to be raised include:

- Should we repeat the Macro view?
- Who needs to know what research data New Zealand Institutes hold and is this more than a storage conversation?
- As a community - what does an end goal look like?

## ABOUT THE AUTHOR(S)

### - Ai-Lin Soo Bio

Ai-Lin Soo has a background in Commerce, with a focus on sustainability and BioMedical Science and has been with the Monash eResearch Centre ([www.monash.edu/researchinfrastructure/ersearch](http://www.monash.edu/researchinfrastructure/ersearch)). Ai-Lin is involved in a number of cross cutting initiatives within the Centre, applying her project management skills to a diverse range of projects from marketing and communications to research data management. Ai-Lin is also heavily involved in the Research Data Culture Conversation ([www.researchdataculture.org](http://www.researchdataculture.org)) and manages the Monash University response to the ARDC Data Retention and Institutional Underpinnings programmes.

### - Claire Rye Bio

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.

- Nick Jones Bio

Nick Jones is NeSI's founding Director, having established and led NeSI alongside a team of colleagues and peers since inception in mid-2011. Nick is responsible for NeSI's partnerships, strategic directions and performance, bringing together a talented and diverse array of people, and their institutions and interests. Nick is a founding member of the Aotearoa New Zealand Committee on Data in Research, and developed the eResearch Ecosystem Map to establish a common framework for eResearch capability and investment nationally and institutionally.

- Rhys Francis Bio

Rhys 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. Since retiring he has developed a revised eResearch Framework for government, assisted the University of Melbourne develop its Petascale Campus Initiative and facilitated the Research Data Culture Conversation ([www.researchdataculture.org](http://www.researchdataculture.org)). Today Rhys continues to contribute to the development of national digital research infrastructures as the Associate Director, Strategy and Partnerships, for the Australian BioCommons ([www.biocommons.org.au](http://www.biocommons.org.au)).

- Luc Betbeder-Matibet

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 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.





# Title: Managing research data at scale: addressing the growing data challenge together

David Jung, UNSW Sydney, [david.jung@unsw.edu.au](mailto:david.jung@unsw.edu.au)

Ai-Lin Soo, UNSW Sydney, [ai\\_lin@unsw.edu.au](mailto:ai_lin@unsw.edu.au)

Adele Haythornthwaite, The University of Sydney

Helena Lynn, The University of Melbourne

Andreas Mertin, University of Technology Sydney

J Max Wilkinson, ARDC

Nichola Burton, ARDC

Rhys Francis, Research Data Culture Conversation, [rhyssfrancis@gmail.com](mailto:rhyssfrancis@gmail.com)

Frankie Stevens, Research Infrastructure Services, [stevens.e.frankie@gmail.com](mailto:stevens.e.frankie@gmail.com)

Kwun Lun (Jacky) Cho, UNSW Sydney

Luc Betbeder-Matibet, UNSW Sydney, [luc@unsw.edu.au](mailto:luc@unsw.edu.au)

## ABSTRACT / INTRODUCTION

Modern institutions face the challenge of uncontrolled expansion and the indefinite retention of uncurated(-able) digital content associated with research activities. Underlying this challenge are two problems: 1) there is low institutional maturity in the ability to assess this digital corpus that prevents institutions and researchers extracting and acting on insights; 2) there is a limited understanding of institutional obligations and best practices around the selection of research content for appropriate disposal and retention. This formed the basis of two ARDC-funded, institutional underpinnings extension projects that commenced in March 2023 (Business Intelligence & Reporting and Retention & Disposal).

Through a series of sector-wide consultation activities, we:

1. Produced a taxonomy that can be applied to digital research content that aids reporting and decision making
2. Developed a minimum research data reporting format and trialled its ability to be reported
3. Unravelling the “data lifecycle” diagram and identified key operational opportunities for metadata enrichment and retention or disposal actions
4. Synthesised best practices at the identified opportunities for compliant and efficient data retention or disposal actions
5. Highlighted sector alignment as a key driver for institutional adoption of project outputs.

In this presentation, the project outputs will be summarised.

## ABOUT THE AUTHOR(S)

- David Jung  
David Jung is a project manager at UNSW and has provided technical, policy, and strategic advice on research data management to researchers and administrators alike in a number of previous roles. He recently led two ARDC-funded institutional underpinnings extension projects that aimed to develop a taxonomy for institutional research data reporting and aligned best practice guidelines for safe and effective research data retention and disposal.
- Ai-Lin Soo  
Ai-Lin Soo has a background in Commerce, with a focus on sustainability and BioMedical Science and has been with the Monash eResearch Centre ([www.monash.edu/researchinfrastructure/ereseach](http://www.monash.edu/researchinfrastructure/ereseach)). Ai-Lin is involved in a number of cross cutting initiatives within the Centre, applying her project management skills to a diverse range of projects from marketing and communications to research data management. Ai-Lin is also heavily involved in the Research Data Culture Conversation ([www.researchdataculture.org](http://www.researchdataculture.org)) and manages the Monash University response to the ARDC Data Retention and Institutional Underpinnings programmes.





# NeSI's Research Developer Cloud: A cloud native platform enabling challenging solutions

*Jun Huh, New Zealand eScience Infrastructure (NeSI), [jun.huh@nesi.org.nz](mailto:jun.huh@nesi.org.nz)  
Kahu Anderson, New Zealand eScience Infrastructure (NeSI), [kahu.anderson@nesi.org.nz](mailto:kahu.anderson@nesi.org.nz)*

## ABSTRACT / INTRODUCTION

In September, 2023, the New Zealand eScience Infrastructure (NeSI) launched its new Research Developer Cloud, built on top of Flexible HPC (Flexi HPC), a new high-performance private cloud platform. The Research Developer Cloud kicked off with a list of Infrastructure-as-a-Service style cloud offerings, also identified as essential building blocks, including but not limited to compute, storage, network, images, and identity services. The Research Developer Cloud supports RSE(Research Software Engineers)/DevOps users to deploy complex and challenging solutions that enable their science needs.

This talk will cover some of the use-cases from our early access tenants, key services and features, alongside a helicopter view of the core technology components, key partnerships and integrations. In particular, a recent use case of NeSI building a training environment will be looked at; with examples using technologies such as Open OnDemand, Kubernetes, and KeyCloak. This talk will also cover DevOps practices applied to the system and examples of interacting with the cloud native platform via code, and how these best practices help develop solutions in ways that are sustainable and reproducible, and also enable collaboration.

Some of the code examples showcased include GitHub actions that trigger automatic build, test, and deploy pipelines; Terraform and Ansible for interacting with the cloud services via code (Infrastructure as Code, IaC) to automate the deployment process; and more.

## ABOUT THE AUTHOR(S)

**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, and data sovereignty related projects in the recent past years, including Aotearoa Genomic Data Repository and Rakeiora Pathfinder project. Recently he has been focused on service design facilitation to help launching of NeSI's new Flexible HPC.

**Kahu Anderson** is a Research DevOps Support Engineer at NeSI with a strong background in Information Technology. He is currently the 'pathfinder' when it comes to NeSI's new Research Developer Cloud, providing the Researcher Support team with information and documentation to facilitate collaboration between researchers and NeSI's new platform.



# NeSI's Flexible HPC - A journey through collaboration

*Blair Bethwaite, New Zealand eScience Infrastructure (NeSI), [blair.bethwaite@nesi.org.nz](mailto:blair.bethwaite@nesi.org.nz)  
Claire Rye, New Zealand eScience Infrastructure (NeSI), [claire.rye@nesi.org.nz](mailto:claire.rye@nesi.org.nz)  
Thomas Berger, New Zealand eScience Infrastructure (NeSI), [thomas.berger@nesi.org.nz](mailto:thomas.berger@nesi.org.nz)  
Jun Huh, New Zealand eScience Infrastructure (NeSI), [jun.huh@nesi.org.nz](mailto:jun.huh@nesi.org.nz)*

## ABSTRACT / INTRODUCTION

In 2021, the New Zealand eScience Infrastructure (NeSI) procured and integrated Flexible HPC (Flexi HPC), a new high-performance private cloud platform. It is envisaged that NeSI's FlexiHPC will provide the sector with a programmable platform for collaboration around science data and support a scalable approach to mid-tier HPC with national expertise complementing localised integration. This talk will cover some of the motivations and use-cases for the new infrastructure by reflecting back on the tenants through the first 3 years, highlighting key partnerships and integrations. There is an ongoing discovery that feeds back into product development, learning across the strong partnership with AgResearch, repositories, sensitive data solutions, training environment, as well as NeSI's own emerging roadmap for migrating its traditional HPC capabilities. All of these examples make flexible use of the different layers of capability; from the underlying data center, network/REANNZ, security, Research Developer Cloud, colocation, and managed services that have been built collaboratively.

NeSI is finding the growing needs to learn and build capabilities around many areas both technical and non-technical; including cybersecurity, cloud native engineering, service design with defining value propositions for customers, building a mature service management approach, and indigenous data sovereignty.

Flexi HPC aspires to play a core role in enabling collaboration with partnering organisations to embrace the shift in the mode of eResearch, in order to address dynamic needs of the researchers, building and supporting growing services and capabilities together with the community, and sharing the knowledge along the way.

## ABOUT THE AUTHOR(S)

**Blair Bethwaite** has worked in distributed computing for over a decade that includes full HPC & cloud systems design, implementation, and operations. Previously at Monash University, Blair most recently led Monash's use of OpenStack to underpin research computing. Originally from Christchurch, in mid-2018 Blair returned to take up the opportunity of becoming NeSI's Solutions Manager, focusing back up the technology stack closer to the user. Blair is leading the work in bringing FlexiHPC, a new high-performance private cloud platform, to NeSI.

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

**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, and data sovereignty related projects in the recent past years, including Aotearoa Genomic Data Repository and Rakeiora Pathfinder project. Recently he has been focused on service design facilitation to help launching of NeSI's new Flexible HPC.

**Dr Claire Rye** is a Product Manager at 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. Most recently, she was a Product Owner for the Ingestion service of the Human Cell Atlas Data Coordination Platform, overseeing the



development of the software infrastructure and metadata standards that supports data sharing across HCA globally, based at the European Bioinformatics Institute (EMBL-EBI).



# Optimizing Research Data Management: A journey into the unknown

Katie Hannan, Rene Tyhouse  
CSIRO  
[katie.hannan@csiro.au](mailto:katie.hannan@csiro.au), [rene.tyhouse@csiro.au](mailto:rene.tyhouse@csiro.au)

## ABSTRACT / INTRODUCTION

In 2021, CSIRO initiated the Science Data Stocktake project with the aim of optimizing research data management processes and aligning with the upcoming Australian Government Data Availability and Transparency legislation. This project focused on the discovery, identification, and documentation of administrative metadata for datasets stored within the Information, Management and Technology (IMT) business unit's storage systems.

Following an intensive 18-month journey involving successive iterations of business requirement refinements, researcher engagements, comprehensive solution architecture documentation, and strategic stakeholder engagement planning, we reached a pivotal milestone - the implementation of a pilot engagement within IMT to evaluate the effectiveness of the Data Stocktake Dashboard, aptly named 'Swordfish.' Swordfish offers a comprehensive overview of data distributed across six distinct physical locations and four storage tiers.

Built on the Starfish file system scanning tool and Microsoft's PowerBI with integrations powered by Hitachi Vantara's Pentaho the end result provides a user experience through the Swordfish dashboard that allows the information to be easily sliced to get visibility into data stored across many unstructured file systems that have evolved over the last 30 years.

Over the course of the last two years, we have meticulously tracked and recorded more than 200 engagement activities with data custodians including researchers and staff who deliver research services. We've been able to identify gaps in service delivery as well as areas for procedural uplift. Subsequently we drafted the intent statement for the CSIRO Data Archive project. This was approved and is now a project within the Managed Data Ecosystem Program.

Thanks to the ARDC Data Retention Project, the Research Data Culture Conversation, and the National Archives of Australia Petabyte-Plus Data Management Special Interest Group we've learned that we're not alone in not knowing what all 40 petabytes of our research data holdings are and that we're heading in the right direction to be better able to understand them.

## ABOUT THE AUTHOR(S)

- Katie Hannan - <https://orcid.org/0000-0002-5689-4133>

Katie is a Research Data Specialist at CSIRO in Adelaide, working with Data Management Systems. She is passionate about storytelling, linking people with information and helping to facilitate learning experiences. Her research interests are in the areas of human computer interaction, digital legacy, and information society.

- Rene Tyhouse – <https://orcid.org/0009-0006-6087-0712>

Rene is an Infrastructure Integration Specialist at CSIRO in Canberra focused on the information lifecycle and the storage systems that underpin it. He strives to optimise the storage offerings through the understanding and aggregation of metrics to be as informed as possible.



# Pawsey: Advancing Australia's eResearch landscape with cutting-edge infrastructure

Mr. Mark Gray  
Ms. Aditi Subramanya  
Pawsey Supercomputing Research Centre  
[Mark.gray@pawsey.org.au](mailto:Mark.gray@pawsey.org.au)  
[Aditi.subramanya@pawsey.org.au](mailto:Aditi.subramanya@pawsey.org.au)

## ABSTRACT / INTRODUCTION

The conference theme provides a fitting platform to delve into the transformative role the Pawsey Supercomputing Research Centre plays in advancing Australian research.

This talk will cover Pawsey's best practices as an eResearch infrastructure, and our learnings along the way in working with our communities and building an accessible and inclusive supercomputing centre.

Through its recent capital refresh, Pawsey has deployed the world's 17<sup>th</sup> most powerful and 4<sup>th</sup> greenest system, Setonix.

The system is designed to deliver high-performance computing with energy efficiency in mind. Setonix is a hybrid system of central processing units (CPU) and graphics processing units (GPU). This heterogeneous architecture enables Setonix to reach significantly higher computer power and prepare for next-generation high-performance computing (HPC) without significantly increasing power draw.

Another important element is Pawsey's warm and cold storage - a 130PB S3 capable multi-tiered storage system, including Ceph object storage and controlled tape storage.

With its strategic partnerships, Pawsey advances eResearch capability through extensive education and training programs, internships, outreach, and research and development.

In the realm of eResearch infrastructure, the evolution and uptake of technology is paramount.

Researchers worldwide are harnessing cutting-edge technologies to enhance their capabilities.

Pawsey, as a leader in the field, continuously innovates by staying abreast of technological trends (such as quantum computing) and translating these advancements into practical solutions for researchers across diverse disciplines.

This approach has led to significant developments in HPC, data storage and management tools, aligning with the evolving eResearch landscape.

Pawsey's commitment to advancing eResearch infrastructure is achieved through a multi-faceted approach.

We embrace HPC technologies that enable researchers to tackle complex problems with unmatched computational power.

Furthermore, we provide researchers with access to comprehensive training, data management and storage solutions that facilitate data-driven research.

Our methods are rooted in collaboration with service providers and researchers, understanding their needs and delivering tailored support.

Pawsey's tangible success is evident in the ground-breaking research enabled by our resources.

Researchers from various domains have utilised our supercomputing facilities and data solutions to accelerate discoveries, from simulating the behaviour of subatomic particles, using AI to predict traumatic brain injuries to modelling the predictions and impacts of climate change.

Our continuous efforts have led to a continuous growth in research publications and recognition, reinforcing the importance of our advanced eResearch infrastructure.

The Pawsey Supercomputing Research Centre stands as a testament to the fruitful synergy between evolving eResearch infrastructure and pioneering research.

Our commitment to staying at the cutting edge of technology, combined with our collaborative approach, empowers researchers to push the boundaries of their fields.

As the eResearch landscape continues to evolve, Pawsey remains a key enabler, ensuring that researchers have the tools and support they need to navigate this exciting and dynamic terrain.



## **ABOUT THE AUTHOR(S)**

- Name: Mark Gray

Mark Gray is the Head of Strategic Partnerships at Pawsey Supercomputing Centre. Mark's early career includes studying at the University of Wisconsin-Madison in atmospheric science and then research software engineering for the MODIS program at NASA Goddard Space Flight Center.

Returning to Australia he worked with several national scientific facilities, building satellite data processing facilities for large-scale data processing. Since 2013 Mark has been advancing Australian research and high-performance computing at Pawsey, initially in DevOps and research software engineering and most recently assuming senior leadership roles. Today Mark is leading opportunity development at Pawsey and is working to make high performance computing more accessible for Australian researchers.

- Name: Aditi Subramanya

Aditi is driven by a passion for uniting communities, solving intricate problems, and forging connections between seemingly disparate concepts.

Aditi's role centres on fostering collaborations that harness supercomputing's potential and bringing together diverse minds to tackle complex scientific challenges. Aditi loves unravelling puzzles and weaving patterns, while advocating for the transformative power of interdisciplinary innovation.





# Practical Advice for Conducting Experimental Research

Rodney Wilson

## ABSTRACT / INTRODUCTION

In this session, Rodney Wilson, Ciena's Chief Technologist of Research Networks, will present the most significant and revealing findings from his 20+ years of experimental research on emerging applications and networking technologies. Mr. Wilson will provide invaluable insights on demo staging, the value of path-finding research, and learning by doing.

## ABOUT THE AUTHOR(S)

Rodney (Rod) Wilson

Chief Technologist, External Research Networks

Rod Wilson serves as Chief Technologist, External Research Networks, where he is responsible for Ciena's leadership & global interactions with universities, government programs and national research and education networks. As part of Ciena's CTO Group, he orchestrates intersections between emerging technologies, advanced architectures and Ciena research and development groups. Mr. Wilson leads technology incubation initiatives and helps drive promising innovations to life. He serves on the ENCQOR Board of Directors and leads Ciena's 5G advanced technology team.

Mr. Wilson received a Bachelor in Electronics Technology from Ryerson Polytechnical Institute in Toronto Ontario. He is a graduate of the Executive Management school at Stanford University in Palo Alto, California, and has received an honorary degree from Algonquin College in Ottawa Canada. He is a recipient of the Queen Elizabeth II Jubilee medal for his service to Canada.



# Progressing Sensitive Research Data Management Maturity at Waipapa Taumata Rau | UoA

Yvette Wharton<sup>1</sup>, Bincy Jacob<sup>1</sup>

<sup>1</sup>Waipapa Taumata Rau | University of Auckland, Tāmaki Makaurau Auckland, Aotearoa New Zealand  
[y.wharton@auckland.ac.nz](mailto:y.wharton@auckland.ac.nz), [bincy.jacob@auckland.ac.nz](mailto:bincy.jacob@auckland.ac.nz)

## ABSTRACT / INTRODUCTION

Based on recommendations from our research data management maturity stocktake for Waipapa Taumata Rau [[link](#)], a Secure Research Environment (SRE) initiative, began in late-2022. An SRE provides a secure remote computing environment for approved researchers to work collaboratively and analyse sensitive data augmented with layers of governance and auditing. The analysed results can be derived or exported, but the original datasets remain unchanged and never leave the SRE. Such monitored systems keep data safe while preventing misuse of data.

This presentation will discuss our journey to uplift sensitive research data management practices and implement a UoA SRE. We will take you through our process for gathering requirements based on the five safes framework, working with a co-design team to assess options, pilot different solutions, understand training requirements, and develop support resources. We will also outline and explore the enablers and the hurdles we have discovered along the way.

## ABOUT THE AUTHOR(S)

**Yvette Wharton** is the eResearch Solutions Lead at the Centre for eResearch, University of Auckland, working on research data management services and researcher enablement projects.  
<http://orcid.org/0000-0002-6689-8840>

**Bincy Jacob** is the Snr eResearch Engagement Specialist working for the RDM programme at Waipapa Taumata Rau. Coming from a background of mass spectrometry and proteomics, she is now working with different communities within university to develop a secure research environment (SRE) for the management of sensitive data.



# Project orchestration and self-service tools

Daniel Bentall  
Plant and Food Research  
[Daniel.Bentall@plantandfood.co.nz](mailto:Daniel.Bentall@plantandfood.co.nz)

## ABSTRACT / INTRODUCTION

Project orchestration, defined as the management of various data sources, pipelines and artifacts in a data project, is a complex and iterative process. A typical machine learning project includes pipelines such as data ingestion, exploratory data analysis, labelling, modelling, and deployment. These actions are rarely executed in a waterfall fashion, and are often evolving, with data, configuration, code, and metrics in constant flux. Tracking these moving parts and understanding their interconnections pose significant challenges.

Effective project orchestration can greatly reduce this complexity and facilitate the creation of basic self-service tools, further simplifying the process for downstream users. This presentation introduces a solution comprised of a project orchestration command-line interface (CLI), DVC- and git-based project repositories and independent pipeline library repositories.

The project orchestration CLI abstracts away DVC and git commands, enabling the creation of reproducible projects composed of generic pipelines and project-specific pipeline configuration. Project repositories manage project data, configuration, code, and artifacts, while the pipeline repositories house collections of related pipelines, such as general data processing pipelines or deep learning pipelines.

The pipeline libraries greatly benefit from using Prefect to define and deploy pipelines. This makes it easy to decouple the software environment and code from the projects which utilise the pipelines, as well as providing powerful python-based pipeline definitions, observability, and infrastructure configuration.

Finally, to demonstrate this solution, we will look at how a user with no programming experience can develop a deep learning computer vision model using a self-service tool built with this system.

## ABOUT THE AUTHOR

Daniel Bentall

I am a data scientist with 5 years of experience in developing deep learning computer vision projects for horticulture and aquaculture at Plant and Food Research (PFR) across a variety of scientific fields. I am currently leading a research aim in PFR's digital twin programme, with a focus on developing information extraction methods from orchard imagery of planar cordon apple trees.



# Software Development Challenges and Best Practices in an Instrument Data Service

Andrew Wilson  
Centre for eResearch, Waipapa Taumata Rau | University of Auckland  
[andrew.wilson@auckland.ac.nz](mailto:andrew.wilson@auckland.ac.nz)

## ABSTRACT

The Centre for eResearch at the University of Auckland is developing an Instrument Data Service, intended to ingest, store and archive data from scientific instruments. I joined this team with the project already underway and spent a significant amount of time familiarizing myself with the code base, technologies and development practices. This talk covers some of the issues I encountered, and how we have begun to address them. Some are specific to academia, but many are applicable to software projects in other settings as well.

In the field of technology generally, staff turnover is often high. It is thus very important to onboard developers quickly and help them to become productive. Code and documentation need to be discoverable, accurate, and follow consistent standards, and systems should give developers the confidence to make changes and know they are unlikely to break something.

In academic settings, developers tend to have more diverse technical and academic backgrounds, with research code being a common entry point. This can bring a valuable diversity of thought, but also a need to bridge the gap between research code and the standards expected in production-grade development. Judicious use of tools and workflows helps to guide and upskill developers to meet these standards.

There are also challenges in applying an agile framework to research-oriented development, including the sometimes disparate responsibilities of team members, and mapping long-running projects to the granularity of sprints.

In this talk I will look at some ways to get the best out of development tools and workflows to address these issues, and discuss some ongoing challenges in this area.

## ABOUT THE AUTHOR

Andrew Wilson is an eResearch Solutions Specialist at Waipapa Taumata Rau | University of Auckland, contributing to the development of a new Instrument Data Service. Prior to joining the university, he worked as a research engineer and software engineer in commercial settings.



# Supporting General Data Protection Regulation for Health Data

Prof Richard Sinnott  
The University of Melbourne  
[rsinnott@unimelb.edu.au](mailto:rsinnott@unimelb.edu.au)

## ABSTRACT

The Australasian Diabetes Data Network (ADDN – [www.addn.org.au](http://www.addn.org.au)) realises the national type-1 diabetes (T1D) registry for both Australia and New Zealand. Unlike other disease registries, ADDN re-uses existing hospital data sets from at present over 50 hospitals. There are over 20,000 T1D patients included in the ADDN registry with over 250,000 hospital visits recorded. The data is uploaded and refreshed twice per year. This data itself covers extensive information for both paediatric and adult patients. This aggregated data is used for many diverse purposes including for example national benchmarking and a diverse collection of “approved” research projects. The ADDN platform is based upon an opt-out based consent process whereby patients (or parents of children) attending hospital clinics are advised of the existence of the registry. At this time they are given the option to opt-out of the registry if they so wish, i.e. to not have their data included in the registry, with no impact on the healthcare they receive. This model has wide spread adoption, ethical approval and has served the needs of ADDN and the clinical research community well. However, the access to and use of personal data is increasingly being challenged with privacy related legislation being introduced to ensure that patients are fully aware and in control of the data that is collected about them and who might have access to it, e.g., for approved research projects. The one time approval that has underpinned ADDN is unlikely to meet increasingly stringent privacy requirements and associated legislation moving forward.

The EU General Data Protection Regulation (GDPR - <https://gdpr-info.eu/>) provides legislation that places the power of the access to and use of data into the hands of the individual (in this case T1D patients). Rather than a single (one off) opt-out consent process, GDPR demands opt-in consent and importantly ongoing agreement for how the data might subsequently be accessed and used. Thus, patients may decide to not allow their data to be used for particular research studies for example. To meet such challenges, ADDN has developed an eConsent mobile app that allows:

- patients to see the data that is collected about them in the registry;
- receive notifications of access requests related to the use of their data for research studies;
- agree/decline access requests related to the use of their data;
- request that their data is removed/deleted from the registry (so called *right-to-be-forgotten*).

This talk will cover the ADDN project and the mobile eConsent app that is currently being developed and rolled out to patients across Australia and New Zealand.

## ABOUT THE AUTHORS

**Professor Richard O. Sinnott** is Professor of Applied Computing Systems and Director of the Melbourne eResearch Group at the University of Melbourne. He has been lead software engineer/architect on an extensive portfolio of national and international projects, with specific focus on those research domains requiring finer-grained access control (security) and those dealing with big data challenges. He has over 450 peer reviewed publications across a range of applied computing research areas.



# Supporting Open Source Leaders in Aotearoa BoF

Organizational Mycology  
[jonah@orgmycology.com](mailto:jonah@orgmycology.com)

## ABSTRACT / INTRODUCTION

New Zealand has been the birthplace of a number of high profile Open Source Software (OSS) projects and many smaller niche projects. This BoF session will be an open conversation around the challenges and opportunities present when creating open source communities in New Zealand and around the world, and what support open source leaders need to be successful.

What resources, support systems, financial structures, talent pathways and other supporting infrastructure could unlock a flourishing of open source projects in and around science in Aotearoa? What could research institutions, businesses and government departments do to encourage more open source approaches? How can the philanthropic sector learn about open source models and support their impact?

In science, OSS projects often start based on addressing a scientific problem, they work to develop a solution in software and share that with a broader community. Many open source developers have a choice to make about whether they want to “go-big”, or “stay-small” as their software gains traction and users. Our research is exploring this decision point, and the ways in which projects all over the world have right-sized their projects to deliver the impacts their communities crave. We’re looking at how projects have succeeded and failed in the past, and working to support new leaders to learn from past mistakes.

After a brief presentation outlining our research efforts. We’ll describe a recently funded project supporting this research, and a cohort based learning, and global Community of Practice development. Together we’ll collaboratively brainstorm past successes and failures in OSS in NZ, uncover what interesting open source efforts are underway today and explore possible futures where we can better support open source leaders in Aotearoa. This will be an engaging, inclusive and open discussion to bring together open source advocates, leaders and supporting institutions to ask:

**“How could we support people to generate a flourishing of Open Source Software by and for Aotearoa?”**

All are welcome who share in a drive to bring open source impacts to the world from the shores of Aotearoa.

*References:*

Understanding Scientific Open Source Software (OSS) Project Leadership –  
<https://orgmycology.com/understanding-oss-project-leadership/>

## ABOUT THE AUTHOR(S)

Jonah has acted as an organizational leader in software and information systems at private sector companies, universities, and international non-profit organizations worldwide. He's worked to sustain the international non-profit The Carpentries, acting as the chief executive of Software Carpentry and Director of Membership at the follow-on organization, The Carpentries, an impact-driven global volunteer coordination organization.

He built innovation spaces, innovative computing environments, and a research informatics team that increased extramural funding competitiveness at The University of Oklahoma. He has worked inside the hyper-growth software company to help advise and structure how skill development and learning are built into the fabric of a globally operating company.

As a consultant, he’s advised and facilitated for [The Veracity Lab](#), [MolSSI](#) (Molecular Science Software Institute), OMSF (Open Molecular Science Foundation) g0v (Gov Zero) Aotearoa, [FutureBees NZ](#), and [Walk Together Aotearoa](#) on their structures, activities, impact, growth, and approaches to sustainment. Jonah received his MS in Forestry and Natural Resources in 2008, and a B.S. in Physics in 2001, both from Purdue University.A



# Supporting the NASA/New Zealand remote-sensing airborne mission: Rongowai

Mike Laverick<sup>1</sup>, Chris Seal<sup>1</sup>, Delwyn Moller<sup>2</sup>, Christopher Ruf<sup>3</sup>, Scott Gleason<sup>4</sup>, Xiaoyou Lin<sup>2,6</sup>, Stephen Musko<sup>3</sup>, Andrew O'Brien<sup>5</sup>, Matthew Wilson<sup>6</sup>, Rajasweta Datta<sup>6</sup>

<sup>1</sup>Centre for eResearch, University Of Auckland New Zealand

<sup>2</sup>Department of Electrical Computer & Software Engineering, University Of Auckland New Zealand

<sup>3</sup>University of Michigan USA

<sup>4</sup>Southwest Research Institute USA

<sup>5</sup>Ohio State University USA

<sup>6</sup>University of Canterbury New Zealand

[mike.laverick@auckland.ac.nz](mailto:mike.laverick@auckland.ac.nz)

## ABSTRACT / INTRODUCTION

This talk discusses our work at the University of Auckland (UoA) to support the payload and science operations for the remote-sensing airborne mission Rongowai (te reo Māori: rongo meaning to sense, wai meaning water) using the Nectar research cloud infrastructure.

Rongowai is a collaboration between NASA, the New Zealand Space Agency, and Air New Zealand to fly a next-generation Global Navigation Satellite System Reflectometry (GNSS-R) receiver on a commercial aircraft to collect climate data. The payload collects reflection data from the surface of Aotearoa, used to measure soil moisture, inundation, flooding, and even coastal erosion.

To support the mission the UoA is facilitating post-flight payload communication, the processing of unstructured binary data into Level-0 and higher science data products (L0+), data hosting, and payload health monitoring.

The payload has been fully supported since first-flight in mid-September 2022 and has experienced several major payload and workflow code updates, in addition to an aircraft lightning strike! Rongowai currently produces a total of ~175GB of binary-through-to L1 products produced per month, with high-level products about to be made available to the public.

## ABOUT THE AUTHOR(S)

Mike is a Senior eResearch Solutions Specialist for the Centre for eResearch at the University of Auckland. Formerly an atomic astrophysicist at KU Leuven, Mike now uses his experience in research and programming to help tackle the ever-growing digital needs of researchers. As part of the Rongowai mission, a collaboration between NASA and the New Zealand Space Agency, Mike has helped develop operational data workflows and visualisation tools. Mike is also a Python aficionado, helping to train and upskill researchers as a Carpentries workshop instructor.



# The Research Software Alliance (ReSA) and NeSI

Georgina Rae, New Zealand eScience Infrastructure (NeSI), [georgina.rae@nesi.org.nz](mailto:georgina.rae@nesi.org.nz)

Nick Jones, New Zealand eScience Infrastructure (NeSI), [nick.jones@nesi.org.nz](mailto:nick.jones@nesi.org.nz)

## ABSTRACT / INTRODUCTION

The Research Software Alliance (ReSA) was established in 2019 with the vision that 'research software and those who develop and maintain it are recognised and valued as fundamental and vital to research worldwide'. Having been involved in ReSA since its formation, in 2023 NeSI joined as an Organisational Member of ResA, expanding Aotearoa's ability to participate in, learn from, and contribute to international efforts advancing the research software ecosystem.

In this Lightning talk we'll share

- an overview of the Research Software Alliance and how Aotearoa has participated in some of their activities to date;
- an introduction to the Amsterdam Declaration on Funding Research Software Sustainability (the ADORE Declaration); and
- information on how to keep up to date with, or become involved in, the kōrero around Research Software

## ABOUT THE AUTHOR(S)

Georgina is the Science Engagement Manager at NeSI where she ensures that NeSI is building strong relationships with the research sector. Prior to NeSI she has worked in molecular biology and intellectual property. She is passionate about enabling research and is interested in the fundamental shifts required to level up scientific research.

Nick Jones is NeSI's founding Director, having established and led NeSI alongside a team of colleagues and peers since inception in mid-2011. Nick is responsible for NeSI's partnerships, strategic directions and performance, bringing together a talented and diverse array of people, and their institutions and interests. Nick is a founding member of the Aotearoa New Zealand Committee on Data in Research, and developed the eResearch Ecosystem Map to establish a common framework for eResearch capability and investment nationally and institutionally.





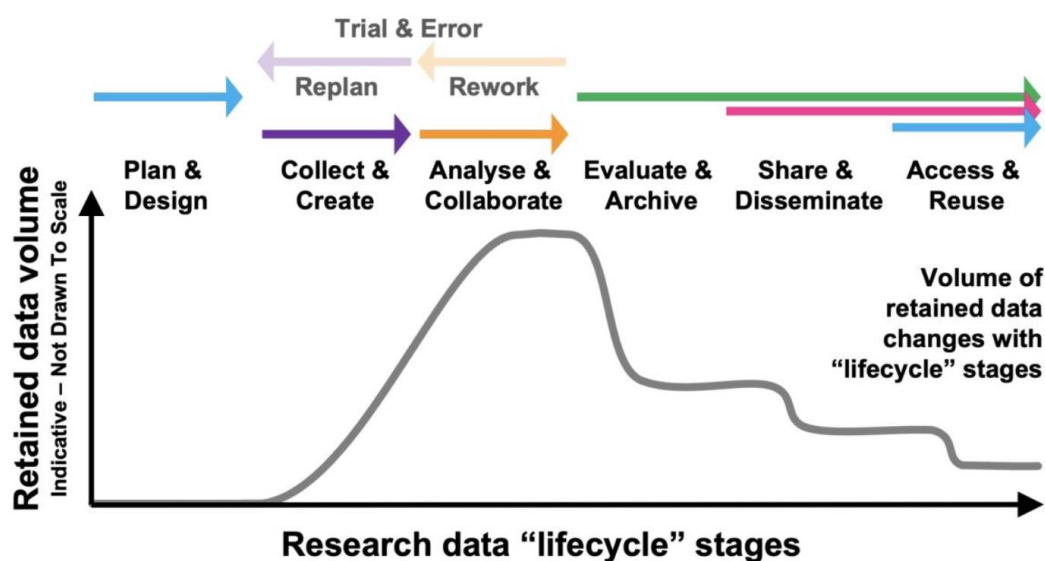
# Unravelling the data lifecycle

Roger Lins, University of Auckland, [rj.lins@auckland.ac.nz](mailto:rj.lins@auckland.ac.nz)  
Rhys Francis, Research Data Culture Conversation, [rhyssfrancis@gmail.com](mailto:rhyssfrancis@gmail.com)  
David Jung, UNSW, [david.jung@unsw.edu.au](mailto:david.jung@unsw.edu.au)  
Ai-Lin Soo, Monash University, [ai-lin.soo@monash.edu](mailto:ai-lin.soo@monash.edu)  
Luc Betbeder-Matibet, UNSW, [luc@unsw.edu.au](mailto:luc@unsw.edu.au)  
Claire Rye, New Zealand eScience Infrastructure, [claire.rye@nesi.org.nz](mailto:claire.rye@nesi.org.nz)  
Nick Jones, New Zealand eScience Infrastructure, [nick.jones@nesi.org.nz](mailto:nick.jones@nesi.org.nz)

## ABSTRACT / INTRODUCTION

(max. length 500 words, excluding Title, Authors and Biography)

Recent work carried out in Australia with New Zealand participation has created a more detailed way to identify the implementation challenges of research data management (a component of the Institutional Underpinnings program of the ARDC). Depicted here as “unravelling” the life cycle diagram, the challenges exposed relate to the nature of the decisions that drive data curation, retention, and disposal in each life cycle stage.



As data volumes rise, it is vital that curation effort focuses on the most important data sets. Noting that these can be expected to be distributed across all New Zealand’s research performing institutions, a commonality of approach will be important.

This BoF aims to bring the community together and focus the discussion and activities around the data culture needed to realise a collective outcome. Brief presentations will be provided of some of the analysis leading to this way of thinking.

Questions to be raised include:

- Data stewardship - what is the goal?
- What should a data steward know about their data?
- How can we support each other to adopt FAIR and CARE principles?

BoF format.

## ABOUT THE AUTHOR(S)

- Name
- Biography of presenting Author (max. length 200 words)
- Roger Lins
- David Jung  
David Jung is a project manager at UNSW and has provided technical, policy, and strategic advice on research data management to researchers and administrators alike in a number of



previous roles. He recently led two ARDC-funded institutional underpinnings extension projects that aimed to develop a taxonomy for institutional research data reporting and aligned best practice guidelines for safe and effective research data retention and disposal.

- Ai-Lin Soo Bio

Ai-Lin Soo has a background in Commerce, with a focus on sustainability and BioMedical Science and has been with the Monash eResearch Centre ([www.monash.edu/researchinfrastructure/eresearch](http://www.monash.edu/researchinfrastructure/eresearch)). Ai-Lin is involved in a number of cross cutting initiatives within the Centre, applying her project management skills to a diverse range of projects from marketing and communications to research data management. Ai-Lin is also heavily involved in the Research Data Culture Conversation ([www.researchdataculture.org](http://www.researchdataculture.org)) and manages the Monash University response to the ARDC Data Retention and Institutional Underpinnings programmes.

- Rhys Francis Bio

- Luc Betbeder-Matibet

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 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.

- Claire Rye Bio

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.

- Nick Jones Bio

Nick Jones is NeSI's founding Director, having established and led NeSI alongside a team of colleagues and peers since inception in mid-2011. Nick is responsible for NeSI's partnerships, strategic directions and performance, bringing together a talented and diverse array of people, and their institutions and interests. Nick is a founding member of the Aotearoa New Zealand Committee on Data in Research, and developed the eResearch Ecosystem Map, a common framework for eResearch capability and investment nationally and institutionally.



# Upskilling researchers in Machine Learning

Mike Laverick<sup>1</sup>, Maxime Rio<sup>2,3</sup>, Ben Collings<sup>1</sup>, Murray Cadzow<sup>4</sup>

<sup>1</sup>Centre for eResearch, University Of Auckland

<sup>2</sup>New Zealand eScience Infrastructure (NeSI)

<sup>3</sup>National Institute of Water and Atmospheric research (NIWA)

<sup>4</sup>University of Otago

mike.laverick@auckland.ac.nz, maxime.rio@nesi.org.nz, ben.collings@auckland.ac.nz, murray.cadzow@otago.ac.nz

## ABSTRACT / INTRODUCTION

Emerging tools and techniques in the space of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) are set to change many aspects of our daily lives and research is no exception. But how do we ensure that researchers are equipped to understand and utilise these tools and techniques across the vast spectrum of research domains?

In this birds-of-a-feather session we aim to survey ongoing efforts around Aotearoa/New Zealand to introduce and train researchers in ML and DL, and aim to communally discuss some of the following topics: developing content and training materials; challenges in teaching and facilitating these techniques; reaching, teaching, and engaging with a broad audience; encouraging the growing community of practice; and providing support for learners and practitioners through compute and other services.

## ABOUT THE AUTHOR(S)

Mike is a solutions specialist for the Centre for eResearch at the University of Auckland. Formerly an atomic astrophysicist at KU Leuven, Mike now uses his experience in research and programming to help tackle the ever-growing digital needs of researchers. As part of the Rongowai mission, a collaboration between NASA and the New Zealand Space Agency, Mike has helped develop operational data workflows and visualisation tools. Mike is also a Python aficionado, helping to train and upskill researchers as a Carpentries workshop instructor.

Maxime Rio is Data Research Software Engineer for NeSI and NIWA. Maxime helps researchers build data science pipelines on NeSI platforms.

Ben is an engagement specialist at the Centre for eResearch at the University of Auckland. With a background in Physical Geography, he stumbled into the world of geographical information systems and remote sensing discovering how useful machine learning could be for resolving landcover classification problems with satellite imagery. He is completing his PhD developing new tools for coastal change detection with satellite imagery. He enjoys learning and teaching to provide researchers with tools to excel in their work and has been involved in several machine learning workshops with CeR and NeSI.

Dr Murray Cadzow: Murray is a Scientific Programmer within Research Teaching IT Support at the University of Otago. Prior to this he spent 11 years researching the genetic basis of gout and related diseases. Murray has been heavily involved in computational literacy and bioinformatic training at the University of Otago - organising Research Bazaar Dunedin and the Otago Bioinformatics Spring School. He is both a Carpentries instructor and instructor trainer. His teaching has focused on delivering digital literacy training to researchers, and the development and support of the local Carpentries community at Otago.



# VUW-SNAP a Network Hub for Simulation, Numerical methods, Analytics and Programming

Patricia Hunt\*<sup>1</sup>, Yvette Perrott<sup>1</sup>, Brendan Harding<sup>1</sup> and Matt Plummer<sup>1</sup>

<sup>1</sup>Victoria University of Wellington, Te Herenga Waka (VUW-THW)

patricia.hunt@vuw.ac.nz, yvette.perrott@vuw.ac.nz, krista.steenbergen@vuw.ac.nz, brendan.harding@vuw.ac.nz and [matt.plummer@vuw.ac.nz](mailto:matt.plummer@vuw.ac.nz)

## ABSTRACT

The computational and data literacy skill sets required by modern academics and researchers across all disciplines have undergone a step change. Numerical modelling, programming, scripting, data analysis and simulation have been brought to the forefront in a modern research environment rich in data and computation. Academics and students without computational expertise are often at a disadvantage. The skill sets desired by many employers are pivoting towards computational expertise and data analysis in addition to the more traditional problem-solving skills. Expert computational researchers at VUW-THW are currently ring-fenced within individual disciplines with infrequent cross-communication despite the commonality of computational methodologies. We are a community-led initiative comprised of academics and e-resource support staff, from a wide range of disciplines and career stages who use and facilitate numerical and computational modelling. We have set up SNAP, a VUW network hub for **S**imulation, **N**umerical methods, **A**nalytics and **P**rogramming with the aim of connecting, supporting and enabling researchers across diverse disciplines, particularly those who have a primary research domain outside of a computing or data-science school. SNAP facilitates the sharing of expertise across the community and acts as a platform for representation at a strategic level. In this presentation I will talk about our journey, the challenges and successes. How did we get started? What were our objectives and have we met them? What does SNAP look like now 18 months on and where do we go from here?

## ABOUT THE AUTHOR(S)

Prof. Patricia Hunt is a Professor of Theoretical and Computational Chemistry in the School of Chemical and Physical Science in VUW-THW. She is the leader of SNAP, heading a committee of SNAP Champions from a wide range of Schools and Research Centres at VUW-THW. Prof. Hunt uses and develops computational methods (quantum and classical mechanics) to study chemical structure and reactions at the molecular level. She recently returned "home to NZ" from her position as a Professor at Imperial College London, UK. She is the leader of SNAP, heading a committee of SNAP Champions from a wide range of Schools and Research Centres at VUW-THW. Dr Perrott and Dr. Harding are or have been deputy leads of SNAP and Matt Plummer is an invaluable contributor to SNAP.



# Weathering the Storm: Engineering Natural Disaster-Resistant eResearch Infrastructure

Aaron Murrhly  
REANNZ  
[aaron.murrhly@reannz.co.nz](mailto:aaron.murrhly@reannz.co.nz)

## ABSTRACT / INTRODUCTION

Aotearoa has a long history of natural disasters from earthquakes and volcanic eruptions to flooding, landslides and wildfires. As we've witnessed in recent years, climate change will only increase the regularity and ferocity of extreme weather events. As many of Aotearoa's research institutions and scientists are involved in studying and responding to such natural phenomena, it is during these events when availability of our research infrastructure is most critical.

This talk will outline the challenges faced, partnerships formed and steps taken by REANNZ to deliver robust research network infrastructure that is resilient to regional disasters. We'll also introduce strategies that can be used by institutions and researchers to leverage existing capability and cutting-edge connectivity technologies to maximise uptime of scientific instrumentation, offices, and campuses empowering kiwi researchers to create a better tomorrow.

## ABOUT THE AUTHOR(S)

- Aaron Murrhly
- Aaron is the Technical Lead for Networks at REANNZ. His passion is for building simple, reliable, scalable networks and advocating for the same within the REANNZ membership. His responsibilities include network architecture and availability, automation and infrastructure security. Outside of work he can usually be found hitting the local trails on his mountain bike or with a ball at his feet on the futsal court.



# What happens after you turn on the flashlight? Building Security Capability

Michael Karich  
University of Auckland / NeSI  
[Michael.karich@auckland.ac.nz](mailto:Michael.karich@auckland.ac.nz)

## ABSTRACT / INTRODUCTION

Over the course of the last 2 years, The University of Auckland has invested in tools, processes, and awareness to broadly build security maturity throughout the university. Through the use of investment in tooling and automation ‘turning on the flashlight’ we have illuminated significant risk and vulnerabilities across many of the services that our researchers use daily. In one service, over 3000 incidents were created for vulnerable researcher systems. We have also identified over 2000 publicly accessible web domains that need to be monitored and secured. Now that we can see into the dark, we are working across teams to build maturity through education, tooling, automation, to sustainably approach this cave of a problem. We have handed researchers the keys to amazing tools and have only now turned on the headlights for them, but across the sector still expect them to be system administrators, security engineers, and application experts on top of completing the great work showcased here. Due to the agile, adaptable, and ever-changing needs of research, it isn’t possible to have IT staff administer these research systems and processes as we do enterprise applications. So we are investigating configurations, tooling, and automation to keep the flashlight on across research systems and patch and configure as broadly and automatically as possible while building in capability to support researchers when it breaks. This is an active work in progress, but we are seeing great traction with our researchers approaching this as collaborators and taking a shared responsibility approach.

## ABOUT THE AUTHOR(S)

Michael Karich is Deputy Chief Information Security Officer – Research at the University of Auckland, NeSI, and UniServices. 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 administration.

With Computer Science and IT Operations qualifications, Michael pulls from a broad background to build and enable secure yet efficient research. He is currently focusing on the enhancement of internationally known capability through governance, staff enablement, and visibility.



# Why we Need a Reference Architecture for Research Data

David Abramson<sup>2</sup>, Luc, Betbeder-Madibet<sup>3</sup>, Stephen Bird<sup>4</sup>, Jake Carrol<sup>2</sup>, Rhys Francis<sup>5</sup>, Wojtek, Goscinski<sup>9</sup>, Ai-Lin Soo<sup>3</sup>, Garry Swan<sup>6</sup>, Carmel Walsh, Glenn Wighwick<sup>8</sup> & J Max Wilkinson<sup>1</sup>.

<sup>1</sup>Australian Research Data Commons, <sup>2</sup>University of Queensland, <sup>3</sup>University of NSW, <sup>4</sup>Queensland Cyber Infrastructure Foundation, <sup>5</sup>University of Melbourne, <sup>6</sup>CSIRO, <sup>8</sup>University of Technology Sydney, <sup>9</sup>National Imaging Facility, [max.wilkinson@ardc.edu.au](mailto:max.wilkinson@ardc.edu.au)

## ABSTRACT / INTRODUCTION

There is little doubt that we have entered an era where data underpins modern science and research in general. In support of this, numerous infrastructures have been designed and built, ranging from proprietary on-premise systems through to distributed commercial clouds. Such implementations provide a range of functions during the research lifecycle from provisioning and cataloguing data assets through to storing and presenting data to computing platforms. In this paper we analyse the underlying principles of such systems and develop a high-level Research Data Reference Architecture (RDRA). Specifically, we identify eight key features of a RDRA that can guide the design, construction, and procurement of implementations without mandating any domain, approach, technical solution, or product choice. As a result, it allows implementers to make local and commercial decisions while still meeting the core requirements of a research data management platform. The intended audience is teams charged with implementing infrastructure in research organisations

## ABOUT THE AUTHOR(S)

- J Max Wilkinson

Max has a comprehensive background in research data management, research data governance and research infrastructure operations. For the last 4 years he has worked with the Australian Research Data Commons as a research data infrastructure architect, designing a scalable and sustainable investment model for nationally significant research data collections. Prior to this, he has worked with the National eScience Infrastructure (NeSI), Council of New Zealand Research Librarians (CONZUL), AgResearch, eResearch2020 and MBIE. He lived and worked in the UK for two decades, most recently as Head Of Research Data and Network Services at University College London, the Datasets Programme Manager at the British Library and Informatics coordinator at Cancer Research UK. He received his PhD in Molecular Nephrology from UCL in 2003.

