eResearch NZ 2023

15-17 February | HAMILTON + ONLINE

Conference Handbook

The eResearch NZ Organising Committee thanks and acknowledges our sponsors:

Silver Sponsors

Bronze Sponsor

Virtual Sponsor

Contents

Table of Contents

Nau mai! Haere mai! Welcome!

Kei āku iti, kei āku rahi, nau mai, whakatau mai ki eResearchNZ 2023!

It is year 14 of eResearch NZ, and our first time back at the University of Waikato since 2014!

After last year's incredibly well attended virtual conference, this year we have returned to a hybrid in-person and online event, and it's fantastic to see both options being embraced by attendees. Although the virtual component adds a layer of complexity from an organisational perspective, it also provides great flexibility for those who are unable to attend in-person, and hopefully makes eResearch NZ accessible to a broader cross-section of the research community.

Over the years this conference has grown to become a key event for New Zealand researchers, providing a forum for national strategic discussions, an opportunity for community building, and a platform for presenting innovative research and teaching applications. Each year has seemed to present us with new challenges, and 2023 looks to be no different.

This year our theme is Capability, Connectivity and Equity, and these simple concepts reflect the multitude of efforts that are currently underway to reshape our research landscape in a way that addresses issues with equity, diversity and opportunity in our research community. This conference is our opportunity to discuss these issues, exchange our ideas, and listen to our friends and colleagues. It is also our chance to share the stories of our work, to learn about new tools and concepts, and to join in planning for the future.

The opportunity to hear from our diverse research community is an important part of this event, and as such, one of our core goals is to provide an inclusive environment where participants are encouraged and supported to actively participate in the event, as outlined in our Code of Conduct.

Thank you to everybody for their contributions, whether as presenters, or through your participation each day. Also, thank you to the programme committee, the conference organising committee, our 2022 partner institution Te Whare Wananga o Waikato University of Waikato, the wonderful University of Auckland Event Services team, Genomics Aotearoa, NeSI and REANNZ as the conference co-hosts, and our very generous sponsors.

It is great to have such strong support from such a wonderful community – I look forward to seeing you all in Hamilton!

Ngā mihi maioha

Mik Black
Chair, eResearch NZ 2023

Committees

Organising Committee

Mik Black (Chair) - Genomics Aotearoa / University of Otago

Rania Alani - REANNZ

Nisha Ghatak - NeSI

Claire Grant - Genomics Aotearoa

Jana Makar - NeSI

Nauman Maqbool - AgResearch

Tyler McInnes - Genomics Aotearoa

Programme Committee

Nauman Maqbool (Chair) - AgResearch

Mik Black - Genomics Aotearoa / University of Otago

Natalie Forsdick - Manaaki Whenua Landcare Research

Maui Hudson - University of Waikato

Leah Kemp - Institute of Environmental Science and Research

Conference Hosts

General Information

If you have any queries not covered here, please visit the registration desk.

Registration desk hours

Conference registration will be in the Gallagher Academy of Performing arts, Calder & Lawson Gallery. It is open during the following hours:

• Wednesday 15th February 09:30 – 16:30
• Thursday 16th February 8:30 – 16:30
• Friday 17th February 8:30 – 16:30

Meals

Morning tea, lunch and afternoon tea will all be served in the Concert Chamber Foyer. All day tea and coffee will be served in the Playhouse foyer. Barista coffee can be purchased from the Opus bar. If you have advised us of any dietary requirements, these considerations will be clearly marked.

Urgent messages and lost property

Urgent messages for delegates and lost property can be directed to the registration desk.

Parking

Parking is available via Gate 2b, Knighton Road, Hamilton, with the Academy a two-minute walk around the campus lake.

Two accessibility car parks are available on the turnaround via Gate 1, Knighton Road.

The parking is free to delegates; please ignore the signs asking for payment; if you do get an infringement notice, please contact the event coordinators to fix this.

Name tags

Delegates are requested to wear their name tags to all sessions and social functions.

Internet

Free Wi-Fi will be available.

Guest WiFi - Open to all delegates/sponsors and does not need home institution authentication –

Wifi code is: Conference
Password: p@ssword

Eduroam - Connect as per instructions from REAANZ on the posters at reception.

Smoking

There is no smoking allowed inside the venue

Presentation Information

Presenting authors

If you are scheduled to give an in-person presentation, please provide your PowerPoint to the AV tech in the Room you are presenting immediately prior to your session. Please have this on a USB.

Files naming convention

Please save your presentation file using the below naming convention to better assist the conference AV tech:

Presentation Session Number__Presenter name (first_ last)

Example: A1_Joe_Smith

Presentations

As a courtesy to our presenters, please ensure you arrive at each session venue prior to the start of presentations.

Mobile phones

During all presentations please switch off or turn your mobile phones to silent.

Timing

Please respect fellow presenters and ensure your presentation does not exceed the given time allocation.

Rooms and AV

Each room features standard audio-visual equipment including a data projector and screen, clicker, presentation laptop and lectern microphone.

Cameras and electronic recording

No electronic recording of presentations is permitted in any form without the express permission of workshop organisers and speakers.

Session Chairs

All Session Chairs – please be in your room at least 10 minutes prior to the start of a session. Please familiarise yourselves with the AV equipment.

If you have any questions, there is an AV technician available. It is crucial that presentations do not run over their allocated time

Getting Around

Public Transport
Timetables and a journey planner can be found at:
https://www.busit.co.nz/

Taxis
Hamilton Taxis – 0800 477 477

Airport Transport
SuperVan – 0800 777 826

Code of Conduct

eResearch NZ, co-hosted by REANNZ, Genomics Aotearoa and NeSI, is an inclusive, harassment free event.

We expect all attendees, speakers, sponsors and volunteers to treat each other with respect, dignity and kindness; regardless of gender or gender identity, sexual orientation or perceived sexual orientation, age, physical ability, appearance, ethnicity, religious or political beliefs, software, hardware and musical preferences.

Harassment, bullying or discrimination will not be tolerated in any form.

If anyone makes you feel uncomfortable or unsafe in any way, or if you see someone else being harassed, please come to the registration desk. All reports and conversations will be taken seriously and handled in confidence.

We value your attendance and are committed to providing a welcoming, safe and equitable environment.

Hamilton Map
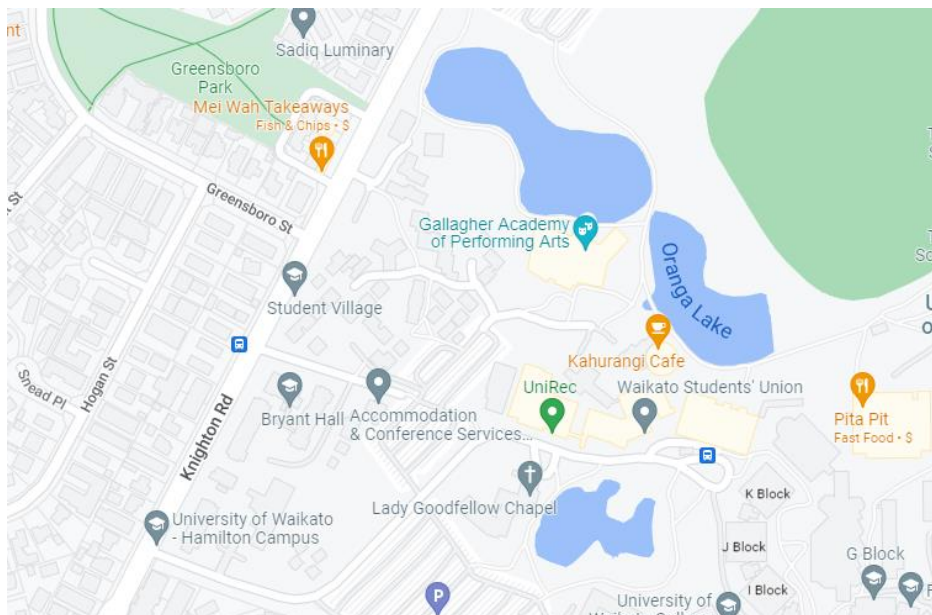
📍 Conference Venue: Gallagher Academy of Performing Arts

Knighton Road University of Waikato, Hamilton 3216

Hamilton Information

New Zealand Emergency Services

Ambulance, Fire and Police. Dial 111 from any public, private or mobile phone in New Zealand.

Hamilton Police
07 858 6200

Hamilton City Hospital
09 367 0000

Pharmacies

Unichem Hamilton Pharmacy. 173 Anglesea Street, Hamilton Central, Hamilton 3204

Open 8am – 8:30am

Medical Centers

Victoria Clinic - GP & Urgent Care, 173 Anglesea Street, Hamilton Central, Hamilton 3204. Open 8am – 8pm

Social Functions

Conference Welcome

Wednesday 15th February 202, 16:30

Gallagher Academy of Performing Arts

Nibbles and refreshments will be provided.

*This event is included in full and student registrations and attendance was reserved during the registration process.

Conference Dinner

Thursday 16th February 2023, 18:00

The Verandah
Rotoroa Drive, Hamilton Lake, Hamilton 3204

Transport to the Dinner

Buses will depart from Ibis Hotel starting at 5:45pm. There will be 3 shuttle rounds.

The buses will return to the Ibis from 9:00pm onwards.

*This event is included in full and student registrations. If you wish to attend the dinner and have not registered, please visit the registration desk.

Your name badge acts as your 'ticket' for this event, please ensure you bring it with you.

View of the Gallagher Academy of Performing Arts

Keynote, 11:00 am, Wednesday 15 February

Data Equity: Addressing the meteoric rise in the importance of Digital Sequence Information

Libby Liggins

Libby Liggins is an Associate Professor at Massey University and a Research Associate of the Auckland Museum Tāmaki Paenga Hira. Libby's research characterises the generation and erosion of biodiversity and she participates in several initiatives interested in the ethical stewardship and aggregation of required data. Libby is on the Scientific Committee of the World Data System, and the Steering Committee for the Genomic Observatories Metadatabase and the Group on Earth Observations Biodiversity Observation Network Genetic Composition Working Group. In Aotearoa, Libby is contributing to the development of the Aotearoa Genomic Data Repository, for which access to the genomic data is controlled by nominated kaitiaki.

Maui Hudson

Maui Hudson affiliates to Whakatōhea, Ngāruahine, and Ngā Puhi. He is an Associate Professor and Director of Te Kotahi Research Institute at the University of Waikato. He is a founding member of Te Mana Raraunga Maori Data Sovereignty Network and Global Indigenous Data Alliance, a co-author of the CARE Principles for Indigenous Data Governance, co-director of Local Contexts, and co-developer of the Biocultural Labels.

Keynote, 9:30am Thursday 16 February

The Silent Genomes Project: Construction of and Indigenous governance over an Indigenous genetic variation reference database for rare disease diagnosis in Canada
Abstract available on page 16

Wyeth Wasserman

Dr. Wasserman is a computational biologist with a research focus on the human genome. His research laboratory based at BC Children's Hospital develops and applies computational methods, software and databases to gain insight in the functional roles of specific DNA sequences in the human genome. His lab assists in the discovery of genetic variation in patient genomes that cause rare genetic diseases. He co-leads the Silent Genomes Project to confront and overcome inequity in reference data for Indigenous children needing genetic diagnosis of rare disorders. From 2013-2021 Dr. Wasserman served as the head of research at BC Children's Hospital, and he remains deeply committed to supporting the growth and success of pediatrics research, including the development of technology to empower patients with access to and control of their digital data. Dr. Wasserman was born and lives on the traditional lands of the Coast Salish peoples, now living and working within the unceded territories of the xʷməθkʷəy̓əm (Musqueam), Sḵwx̱wú7mesh (Squamish) and Səl̓ílwitulh (Tsleil-Waututh) Nations. He is currently enjoying a sabbatical at the University of Otago.

Keynote, 9:30am Friday 17 February

" Ka Kimi Mana Taurite i te Ao Matihiko" – "Finding equity in the digital world."
Abstract available on page 17
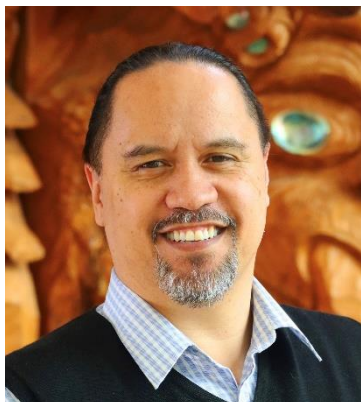
Warren Williams

Dr Williams has extensive experience in IT, business and tertiary education spanning more than 25 years. He has held executive and senior management roles and is active in tribal and community-level groups. As the Chief Executive for 20/20 Trust, he is passionate about enabling and empowering people through education and technology. Digital inclusion programmes are the core of the 20/20 Trust service provision.

Programme – Day 1

11:00	**REGISTRATION OPEN**
12:00- 12:30	**LUNCH**
	CONFERENCE OPENING
	Location: Dr John Gallagher Concert Chamber
12:30	Opening
12:50	John Creech - A discussion of some MBIE initiatives for Aotearoa New Zealand's RSI system *VIRTUAL PRESENTATION*
13:20	**MINI BREAK**

	SESSION A			
	Session A1 **Location: Dr John Gallagher Concert Chamber**	**Session A2** **Location: Te Whare Tapere Iti**	**Session A3** **Location: Playhouse Theatre**	**Workshop** **Location: Upstairs Lounge**
13:30	Max Wilkinson: Building a Coherent and Consistent National Research Data Storage Investment Model	Annie West: Utilising Aotearoa's computing capability in microbial ecology: from kākāpō to estuaries	David Fellinger: Recent Developments and Partnerships in iRODS Technology Enable Global eResearch Collaboration	Committee on Data in Research / Michelle Blake: The state of data in research within Aotearoa *In Person Only*
13:50	Anton Angelo: Why do researchers say they'll share their data, and then don't?	Alvaro Orsi: From Reactive to Proactive: Enhancing Aotearoa's Wastewater Surveillance Programme with Data Science	Alexander Pletzer: Common workflows patterns in scientific computing	
14:10	**MINI BREAK**			
14:20	Ryan Perry: A data risk assessment and education tool for the Australian Data Archive using R Shiny	Joseph Guhlin: Reannotating 282 Genomes: The Hymenopteran Unified Gene Set *VIRTUAL PRESENTATION*	Vladimir Mencl: Enabling network connectivity with eduroam Visitor Access	
14:40	Claire Rye / Greg Darcy : Accelerating data movement between New Zealand and Australia	Guilherme Cassales: Time-evolving data science applied to forest hydrology data	Martin Feller: Security vulnerability management in the Nectar research cloud at the University of Auckland *VIRTUAL PRESENTATION*	
15:00 - 15:30	**AFTERNOON TEA**			

	SESSION B			Workshop continued
	BOF **Location: Dr John Gallagher Concert Chamber**	**BOF** **Location: Te Whare Tapere Iti**	**BOF** **Location: Playhouse Theatre**	
15:30	Libby Liggins / Maui Hudson: What Māori metadata can and should be retained alongside data? *IN PERSON INTENDED*	Yvette Wharton: Institutional approaches to sensitive research data *IN PERSON INTENDED*	Nisha Ghatak: Building Research Capabilities With The Carpentries *IN PERSON INTENDED*	
17:00	**WELCOME FUNCTION**			
	Gallagher Academy of Performing Arts			

Programme – Day 2

Time	Session C1 / D1 / BOF (Concert Chamber)	Session C2 / D2 / BOF (Te Whare Tapere Iti)	BOF (Playhouse Theatre)	Workshop (Upstairs Lounge)
8:30	REGISTRATION OPEN			
	HOUSEKEEPING AND KEYNOTE			
	Location: Dr John Gallagher Concert Chamber			
9:15	Housekeeping			
9:30	Wyeth Wasserman: The Silent Genomes Project: Construction of and Indigenous governance over an Indigenous genetic variation reference database for rare disease diagnosis in Canada			
10:30-11:00	MORNING TEA			
	SESSION C			
	Session C1 — **Location: Dr John Gallagher Concert Chamber**	**Session C2** — **Location: Te Whare Tapere Iti**	**BOF** — **Location: Playhouse Theatre**	**Workshop** — **Location: Upstairs Lounge**
11:00	Nick Jones: eResearch, NeSI and beyond	Kevan Cote: Classifying animal behaviour using machine learning: A collaborative approach to build capability across disciplines	F Jorge Bornemann: Building communities of practice *VIRTUAL BOF*	Shivonne Londt and Karl de Borst: Learn how to accelerate your research with Machine Learning on AWS *IN PERSON ONLY*
11:20	Georgina Rae: Taking an Agile Service Management Perspective at NeSI	Ben Halliday: Detection and Mitigation of Microbiome DNA Presence in Saliva-Derived Whole Genome Sequence Data		
11:40	MINI BREAK			
11:50	Dr Nauman Maqbool: Developing eResearch Services - An update on AgResearch and NeSI partnership on building national capability	Zoe Yang: Pangenome graph: towards a "reference freer future"		
12:10	Blair Bethwaite: AgResearch's eResearch Infrastructure - a capability development journey with NeSI	Chris Scott: War on Wildings: continuing Flow Transport Model development		
12:30 - 13:30	LUNCH			
	SESSION D			Workshop continued
	Session D1 — **Location: Dr John Gallagher Concert Chamber**	**Session D2** — **Location: Te Whare Tapere Iti**	**BOF** — **Location: Playhouse Theatre**	
13:30	Marion Shadbolt: Advancing Human Genomics data sharing in Australia: Highlights from Australian BioCommons' Human Genome Informatics initiative	Richard Dean: How will the GPU in your pocket change the future of eResearch?	Tom Saunders: ResBaz Aotearoa: Building on Success *IN PERSON ONLY*	
13:50	Libby Li: Data Stewardship Wizard: A tool for Digital Data Management Plans	Maxime Rio: Hyperparameter optimisation for Machine Learning: A little bit of theory and practice		
14:10	MINI BREAK			
14:20	Richard Hartshorn: Supporting a national data community - the Committee on Data in Research (CoDiR)	Jonny Williams: Using HPC-driven climate simulations and open-source software to predict future tropical cyclones		
14:40	Andrew McKenna-Foster: FAIR Capabilities: Quantifying the benefits of institutional repositories for graduate students' research outputs	Matthew Adlam: The dynamics of Dictystelium discoideum's microbiome and genomic changes of the symbiont Paraburkholderia agricolaris		
15:00-15:30	AFTERNOON TEA			
	SESSION E			
	BOF — **Location: Dr John Gallagher Concert Chamber**	**BOF** — **Location: Te Whare Tapere Iti**	**BOF** — **Location: Playhouse Theatre**	**Workshop** — **Location: Upstairs Lounge**
15:30	Nooriyah Lohani: Is your institution supporting FAIR research software? *IN PERSON ONLY*	Jana Makar: Building diverse teams through inclusive hiring practices *IN PERSON ONLY*	Blair Bethwaite: Cloud use-cases in research *IN PERSON ONLY*	Claire Rye: Can we characterise Aotearoa New Zealand's research data at scale *IN PERSON ONLY*
18:00	**Conference Dinner**			
	The Verandah Café and Events Centre			

Programme – Day 3

8:30	**REGISTRATION OPEN**				
	HOUSKEEPING AND KEYNOTE				
	Location: Dr John Gallagher Concert Chamber				
9:15	Housekeeping				
9:30	Warren Williams: "Ka Kimi Mana Taurite i te Ao Matihiko" - "Finding equity in the digital world"				
10:30-11:00	**MORNING TEA**				
	Concurrent Session F				
	PANEL DISCUSSION Location: Dr John Gallagher Concert Chamber			**BOF** Location: Playhouse Theatre	**Workshop** Location: Upstairs Lounge
11:00	Panel: Enabling Māori, Pasifika, other minorities' Panel			Yvette Wharton: An unexpected journey into PIDs - from planning to impact and back again *IN PERSON ONLY*	Callum Walley: Working with Singularity Containers (Interactive Workshop) *Available in Person and Online*
12:30 - 13:30	**LUNCH**				
	Concurrent Session G				Workshop continued
	Session G1 Location: Dr John Gallagher Concert Chamber	**Session G2** Location: Te Whare Tapere Iti	**BOF** Location: Playhouse Theatre		
13:30	Mik Black: Building a taonga species data repository for Aotearoa New Zealand	Vladimir Mencl: Improving Trust in Identity Federations: an Update on New Developments in Tuakiri	Laura Armstrong: Incentivising Open Research *IN PERSON ONLY*		
13:50	Ben Te Aika / Claire Rye: Rakeiora a pathfinder for genomic medicine: Building a prototype Genomics Platform	Jun Huh: Flexi HPC - Enabling new ways of eResearch			
14:10	**MINI BREAK**				
14:20	Natalie Forsdick: Journeying towards good data management practices: Tips and tricks to empower the biodiversity genomics community	Sam Hitchman: Automated CT analysis to identify low emitting ruminants *LIGHTNING TALK*			
		Murray Cadzow: Testing Spatial Navigation with Google Street View *LIGHTNING TALK*			
14:40	Paula Andrea Martinez: The first RSE Asia Australia Unconference *VIRTUAL PRESENTATION*	François Bissey: Getting to know what runs in my private research cloud *VIRTUAL PRESENTATION*			
15:00	**CONFERENCE CLOSE**				

Abstracts

The Silent Genomes Project: Construction of and Indigenous governance over an Indigenous genetic variation reference database for rare disease diagnosis in Canada

Dr. Wyeth W. Wasserman
Professor of Medical Genetics, University of British Columbia, Vancouver BC, Canada.

ABSTRACT

The Silent Genomes Project aims to establish Indigenous governance for and creation of a reference genetic variation database to support the equitable diagnosis of the causes of rare disorders for Indigenous patients in Canada. At present there are inequities in the capacity to diagnose Indigenous children in Canada with rare diseases. It can be unclear if observed genetic differences between an Indigenous patient and international reference databases of genetic variation reflect an absence of Indigenous participants in the reference data or a genetic cause of a patient's disease. To address this inequity, the Silent Genomes Project pursues three technical aims shaped by an encompassing aim of Indigenous data and research governance. The three technical aims include: (Aim 2) the use of whole genome sequencing to diagnose Indigenous patients with rare disorders; (Aim 3) the creation of a reference Indigenous Background Variant Library (IBVL) reporting the frequency of genetic variants across the human genome for Indigenous participants; and (Aim 4) a health economics assessment of the impact (under multiple measures) of reference data on diagnosis. Aim 1 focuses on the development of Indigenous governance over the research and the data, which wraps around and within all other aims. The presentation will provide an overview of the Indigenous peoples of Canada, and will focus on Aims 1 and 3, highlighting how bioinformatics can be enabling to the process and implementation of governance models. The similarities and differences between the Silent Genomes Project and the Genomics Aotearoa Variome Project will be explored.

"Ka Kimi Mana Taurite i te Ao Matihiko"
"Finding equity in the digital world"

Dr. Warren Williams
warren.williams@2020.org.nz

ABSTRACT

Please include any of the following points, but not limited to:
This presentation will highlight the building of research capability, connectivity and equity within a Māori context. It will share experiences and aspirations that relate to Māori communities, research and Mātauranga Māori.

- The themes from a iwi/hapū context for Mātauranga Māori and eResearch.
- The opportunities and challenges of connectivity and access to technology for Māori communities.
- How to foster partnerships and relationships with Māori, industry, subject experts, and government that align with authentic Tangata Whenua and Tangata Tiriti engagement and Te Tiriti o Waitangi.
- Thoughts and reflections about the 'Digital Strategy for Aotearoa' released in Sept 2022 and what does/could this mean for Māori.

ABOUT THE AUTHOR

Dr Williams has been in the ICT business and tertiary education industry for over 25 years. He is currently the CEO of 20/20 Trust that provides access and connection to digital skills and resources that empower and enable people to realise their aspirations.

His governance and leadership experience includes national and regional boards, working with Māori and entrepreneurship groups, tertiary ICT advisory groups, and technology and innovation groups. He is passionate about Strategic Māori leadership and succession planning in decision-making forums, especially with rangatahi Māori and wāhine Māori.

A discussion of some MBIE initiatives for Aotearoa New Zealand's RSI system

John Creech
Ministry of Business, Innovation and Employment
john.creech@mbie.govt.nz

ABSTRACT
The Ministry of Business, Innovation and Employment (MBIE) has a role as both a funder of research in Aotearoa New Zealand and as a kaitiaki of the Research, Science and Innovation (RSI) system. This presentation is an opportunity to discuss some of the ongoing work at MBIE across those roles, which aim to benefit researchers, the RSI system, and Aotearoa New Zealand as a whole. Topics may include

- the forthcoming Kitmap online tool to connect researchers with the infrastructure they need to carry out their research
- opportunities for RSI data stewardship around best practices, including Open Science, research data management and Māori data sovereignty
- ways to create a more coordinated, system-wide approach to research infrastructure
- efforts to better understand and increase the impact of our research activities-
- how the above fit within the larger system reforms of Te Ara Paerangi – Future Pathways

ABOUT THE AUTHOR
John is a Senior Policy Advisor in the Science Policy team at MBIE. His work focus spans the data side of the RSI system, including Open Science, Māori data sovereignty, research metrics, impact of research, and research infrastructure. John has a background in planetary science, having previously worked in isotope geochemistry and cosmochemistry research in New Zealand, Europe and Australia.

Building a Coherent and Consistent National Research Data Storage Investment Model

Dr J Max Wilkinson, Ms Carmel Walsh
Australian Research Data Commons
max.wilkinson@ardc.edu.au, carmel.walsh@ardc.edu.au

ABSTRACT

The Australian Research Data Commons (ARDC) believes that to maximise the impact of research data output of meritorious research, researchers must have timely access to high quality data collections stored on stable and persistent infrastructure. The ARDC Data Retention Project (2020-2023) was designed to meet this challenge by understanding the nature and dynamics of the national data storage sector, then tying a coherent investment model to measures of impact and benefit for researchers, eResearch services providers and strategic priorities. Our approach established a standardised, international metadata specification to identify and characterise data collections of national significance then use evidence-based business intelligence to calculate consistent subsidy options.

Our project identified targets where investment impact could be maximised; examples include tools to incentivise metadata management, flexibility in subsidy calculations, assigning incentives to sector stakeholders, amongst others. The project invested in short innovation proofs-of-concept and leveraged ongoing work by self-forming sector partners (particularly the G05 'Research Data Culture Conversation', $10.5281/zenodo.3887433$) that identified similar issues on the scale and dynamics of national data storage sufficiency. Finally, we are establishing a formal definition of 'Data Collections of National Significance', articulating the consequences of any such designation.

In this final year of the Data Retention Project, outcomes and findings will be combined to build a coherent and consistent national investment model for Australian research data storage infrastructure that supports a robust and enduring national data commons. This talk will present the principles and key components of an enduring and sustainable investment strategy.

ABOUT THE AUTHORS

Dr J Max Wilkinson has a comprehensive background in research data management, research data governance and research infrastructure operations. For the last 3 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.

Ms Carmel Walsh is the Director, eResearch Infrastructure & Services at the Australian Research Data Commons (ARDC). The ARDC provides Australian researchers with competitive advantage through data. It does this by enabling the Australian research community and industry access to nationally significant, data intensive digital research infrastructure, platforms, skills and collections of high-quality data. Carmel leads on storage and compute with a focus on the national research cloud compute service for Australian researchers, the ARDC Nectar Research Cloud, and national Data Retention project.

Why do researchers say they'll share their data, and then don't?

Anton Angelo
University of Canterbury
Research, Systems and Data Librarian

ABSTRACT

The list of journals and publishers that require data sharing is growing. Unfortunately experience and research shows that many researchers, though they say their data will be available refuse to provide it. The literature shows a number of reasons this is, and points to some things that those supporting research can do to help, to improve transparency, re-use, openness and integrity.

This 15-minute presentation will outline the current research, and draw themes and practical suggestions to increase the amount of data that is made available from research.

ABOUT THE AUTHORS

Anton has been helping researchers with Data at the University for the last 10 years. He is a qualified Carpentries instructor. His own research interests are in the impact of theses and dissertations.
https://orcid.org/0000-0002-2265-1299

A data risk assessment and education tool for the Australian Data Archive using R Shiny

Ryan Perry, Janet McDougall, Weifan Jiang, Tina Gregor
Australian National University
ryan.perry@anu.edu.au, janet.mcdougall@anu.edu.au, weifan.jiang@anu.edu.au, tina.gregor@anu.edu.au

ABSTRACT

A challenge for public data repositories is ensuring both the public availability of data and the privacy of research participants. At the Australian Data Archive, archivists are required to guide data owners through a data risk assessment and deidentification process to ensure compliance with privacy legislation and with internal procedures for handling sensitive data.

Having developed code in R for this purpose, we aim to make this code user-friendly and publicly accessible via a web-based R Shiny tool. The R Shiny data app we are developing - the ADA DRAT (Data Risk Assessment Tool) - will guide users to assess their own data on a series of privacy criteria in preparation for submission to the archive. The DRAT automatically identifies variables with potentially identifying information, including by combining variables to detect unique combinations of response categories that might lead to identification of individual respondents. Researchers can then examine different options for recoding their data and observe corresponding reductions in identification risk.

In this presentation we will describe archiving procedures at the Australian Data Archive and outline challenges working with data owners to identify privacy risks and appropriately deidentify their data. We will demonstrate current capabilities of the DRAT that can automate risk detection and deidentification, including new functions for handling longitudinal and panel data. We will also discuss preparations for deployment of the DRAT as a self-service tool including educational features, and the generation of data privacy reports.

ABOUT THE AUTHORS

Ryan Perry is a data archivist at the Australian Data Archive. He specialises in archiving national, longitudinal health and economic data. His research examines the role of personality and ideology on political affiliation and policy support. He completed his PhD at the University of Auckland, New Zealand in 2014.

Janet McDougall is a Senior Data Archivist with the Australian Data Archive, ANU. Her background includes mainframe IT, data management, GIS, and social research (ANU MSR Adv). Janet specialises in archival workflows and sensitive data – curation, preservation, technical implementation, controlled vocabularies, and metadata standards.

Tina Gregor is a data archivist at the Australian Data Archive. Tina completed her PhD at Australian National University in 2021 and has previously worked as an archivist at PARADISEC. She specialises in archiving audio/visual and qualitative data.

Weifan Jiang is a data archivist at the Australian Data Archive. Weifan has completed a Master of Machine Learning and Computer Vision at Australian National University. He and has considerable experience using R and Python, and his previous work has included analysis and visualisation of massive Meteorological and Digital Elevation Data at CSIRO.

Accelerating data movement between New Zealand and Australia

Claire Rye, New Zealand eScience Infrastructure, University of Auckland,
Ryan Fraser, Aarnet, Perth, Australia,
Greg D'Arcy, Chris Myers, Aarnet, Melbourne, Australia,
Richard Tumaliuan, Research Education Advanced Network New Zealand (REANNZ),
claire.rye@nesi.org.nz, ryan.fraser@aarnet.edu.au, greg.darcy@aarnet.edu.au, chris.myers@aarnet.edu.au,
richard.tumaliuan@reannz.co.nz

ABSTRACT

In order to lower barriers and to normalise moving substantial volumes of data to enable data intensive science in New Zealand and Australia; a uniform, parallel and secure data transfer and management service is needed. Too often, researchers have been resorting to using portable storage devices to move data around as transfers fail. A democratised service to support data movement, leveraging the network was needed.

Both New Zealand and Australia (through NeSI/REANNZ and AARNet respectively) have adopted Globus. Globus is a service that leverages high-bandwidth networks to provide fast, parallel and secure file transfers and data management; has proved to be well equipped to support researchers and importantly research infrastructure service providers in meeting their transfer and data movement needs.

New Zealand has utilised Globus since 2014, this expanded in 2019, launching a new and improved National Data Transfer Platform in partnership with Globus and the national research and education network REANNZ via science DMZ and are driving adoption widely across the research system.

The deployment of Globus across Australia started in 2021, and has seen good uptake so far, making the most of the extremely high-bandwidth network capacity that is available within (and out of) Australia for the research and education sector.

We will present the early results and learnings from trans-Tasman data transfers, along with future-plans for rollout across Australia and New Zealand to support growing data needs of researchers. Further, we will highlight the opportunities for the sector be that researchers, infrastructure and service providers through the utilisation of this service.

ABOUT THE AUTHORS

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).

Ryan Fraser has had a career working with many amazing researchers, developers and analysts in multiple sectors delivering innovative solutions to support industry and research programs. Ryan currently has the privilege to continue his career at AARNet and support an amazing team delivering eInfrastructure for the Australian R&E sector. Ryan is the Director, Digital Research at AARNet.

Greg D'Arcy is the Research Engagement Strategist, AARNet. Greg is an experienced Research Analyst and Program Manager working across a number of product development projects and research collaborations at AARNet. Greg has over twenty years' experience in the research and education sectors working across several large-scale digital transformation and infrastructure initiatives.

Chris Myers is a DATA Solutions Specialist at AARNet (Australian Academic and Research Network). Chris has 28 years of experience working in the telecommunications and higher education sector especially in large data projects. Chris supports a wide array of data movement projects in training, consultation and design in the areas of astro/high energy physics, HPC, scientific Instruments, archives and film, cloud solutions, Data Centres fabrics and cybersecurity. Chris is participating in the ACCS WP4 project and is lead for WP4.3.

Richard Tumaliuan is a Senior Network Engineer, Research and Education Advanced Network New Zealand (REANNZ). Richard has had over a decade of experience working in the enterprise, government, mining and telecoms industry. Currently part of an amazing team at REANNZ that supports the research and education sector in New Zealand where one of the motivations is to develop and deliver solutions to support the growing needs of the community to transfer large amounts of data on a domestic and global scale.

Utilising Aotearoa's computing capability in microbial ecology: from kākāpō to estuaries

Annie G. West

Waipapa Taumata Rau – The University of Auckland

annie.west@auckland.ac.nz

ABSTRACT

Microbial ecology research in Aotearoa New Zealand spans a variety of flora, fauna and natural environments. Historically, microbiology research was heavily based on cultivation and microscopy techniques. We can now obtain multitudes more information using a range of DNA sequencing techniques, from gene amplicon-based sequencing through to meta-omic technologies. This presentation will give an overview of how both approaches can be leveraged to improve our understanding of the microbial communities around us, from the gut microbiomes of critically endangered species to communities dominating our local estuaries. Among the most important factors underpinning the studies in question is our ability to utilise NeSI resources for processing and analysing these data, as well as the associated community who provide valuable help and feedback in all areas of our research.

ABOUT THE AUTHOR

Annie G. West is a Genomics Aotearoa Postdoctoral Research Fellow working with Dr Kim Handley at Waipapa Taumata Rau - The University of Auckland. Annie completed her PhD at the same institution under the supervision of Professor Mike Taylor on the microbiome in threatened species conservation. Her research on kākāpō and takahē revealed key information regarding the impact of conservation management on the gut microbiomes of both species that has important implications for future management practice (some of which have already been implemented). She now investigates strain heterogeneity and salt tolerance strategies of abundant bacteria, particularly Actinobacteria, across the Waiwera estuary as part of the Genomics Aotearoa (GA) Environmental Metagenomics project. A key facet of Annie's research is to develop user-friendly bioinformatics pipelines for the assessment of strain heterogeneity in metagenomics data that are available to the wider community. She is an Executive Member of the New Zealand Microbiology Society and co-chair of the Genomics Aotearoa Postdocs Group. You can find her on Twitter as @anniegwest.

From Reactive to Proactive: Enhancing Aotearoa's Wastewater Surveillance Programme with Data Science

Lillian Lu (Presented by Alvaro Orsi)
Institute of Environmental Science and Research (ESR)

ABSTRACT

ESR has been at the forefront of wastewater surveillance since the early days of the COVID-19 pandemic. The programme started with collecting and analysing samples, and producing weekly reports to share insights with the Ministry of Health.

In February 2022, rapid antigen test (RAT) was introduced to New Zealand and soon replaced the traditional PCR method as the primary method for COVID-19 diagnoses. Many people are concerned that the self-reported case numbers may not reflect the true level of COVID-19 transmission risks. Luckily, wastewater can help. Our team turned the huge volume of wastewater data ESR possess into a mobile-friendly dashboard, which allows the public to assess the COVID-19 situation in their local areas in a timely manner.

To continue to enhance the wastewater surveillance programme and prepare New Zealanders in a proactive position for dealing with infectious diseases, our team applies machine learning and AI to the wastewater data and aims to provide reliable forecasting results that can be used to support government decision-making. Specifically, we are investigating and comparing the different time series models offered by the Python library Darts, ranging from Regression and ARIMA to RNN and TFT.

During this talk, Lillian will share how the team incorporated data science into the wastewater surveillance programme and brought new values to the science work. She will discuss the secrets of a successful public dashboard (which lies in *connectivity* and *equity*), and the challenges the team encountered while building up their AI and e-research *capability* in the space of infectious disease modelling.

ABOUT THE AUTHOR

Lillian Lu is a data scientist at ESR. She works across the forensic, health and environment science groups on different projects including dashboard development, statistical analysis and machine learning. Lillian co-led the development of the Wastewater Surveillance Mobile Dashboard and is now focused on the relevant modelling work. With a master's degree in analytics and a background in marketing, Lillian is passionate about enhancing ESR's science through analytics applications and public communications.

Reannotating 282 Genomes: The Hymenopteran Unified Gene Set

Joseph Guhlin1,2, Phoebe Keddell2, Peter Dearden1,2
1 Genomics Aotearoa, Department of Biochemistry, University of Otago
2 Department of Biochemistry, University of Otago
joseph.guhlin@gmail.com phoebe.keddell@gmail.com peter.dearden@otago.ac.nz

ABSTRACT

The Hymenopteran lineage is essential for pollination, pest control via parasitism, economically damaging as pests, and scientifically fascinating, partly due to spanning the eusocial gradient. Our comparative genomics paper of three invasive wasps shows that we predicted more genes than expected compared to the greater Hymenopteran gene set. Yet, nearly all of them were well conserved. Lineage-specific genes are often a result of differences in annotation methodology rather than a kink in evolution (Weisman, 2022). To better study evolution and specific classes of genes, such as those linked to eusociality, odorant receptors, and parasitism, we must annotate genomes to remove biases from multiple methods. To do this, we developed a graph-based comparative approach, taking all possible open-reading frames from all genomes and creating a genomic position-aware pan- orf-ome. Similar putative exons are clustered and aligned. ORFs and their relationships are trimmed as more genomes are added. Exon boundaries are identified using these alignments, pruning of extraneous amino acids, and via the AUGUSTUS Comparative Genome Pipeline. By using fast databases, both graph and byte storage, as well as taking advantage of job schedulers, we can complete this once insurmountable task much faster than previously. We can find other interesting genomic features by looking for conserved ORFs rather than simple gene prediction. We use the graph nature of the database to look for exon switching and expand the software to look for pseudogenes. Additionally, we run transposable element finders and lncRNA finders. In the end, we are left with two crucial outputs: a fast comparative genome annotation pipeline and, importantly, a gene set for an entire lineage which has been called with the same methods, based on sequence conservation, trained with said data to find lineage-specific genes, and thus, a set of high- quality genes for 282 genomes which is primed for use in phylogenomic analyses.

ABOUT THE AUTHORS

Joseph Guhlin is a Postdoctoral Research with Genomics Aotearoa at the University of Otago Biochemistry Department in Peter Dearden's lab. Joseph works on population genomics, including the recent Kākāpō population pre-print and high-quality software tools, and is an avid Python and Rust programmer. Other interests include multi-omics (Omics Database Generator), high-quality assemblies, and high-quality annotation.

Time-evolving data science applied to forest hydrology data

Guilherme Cassales1, Justin Liu1, Dean Meason2, Albert Bifet1
AI Institute1, SCION research2
guilherme.cassales@waikato.ac.nz
justin.l@waikato.ac.nz
abifet@waikato.ac.nz
Dean.Meason@scionresearch.com

ABSTRACT

Data are the component that sets the world in motion nowadays, and they are generated and used at ever-increasing rates. As a result, their value has skyrocketed thanks to the capability of helping decision-making through the data-driven approach.

Environmental science is one of the prominent topics among the many applications that stem from the data analytics field because of its relevance in our society and the amount of data it generates. In such a field, data are essential to research, understand, set policies, and manage the environment. For example, to maintain ecological balance in planted forests, one must understand how water flows through the land to make the best use of land and water. In such cases, the amount and complexity of the environmental data can present many challenges to extracting meaningful knowledge, creating the demand for new research to overcome these challenges and replace outdated models. Therefore, technical innovation and scalable systems that leverage the use of data science are required to improve the interpretation capability of environmental researchers, governors, and managers.

The forest hydrology problem presented in this work is a collaboration between the Forest Flows and TAIAO projects. From the Forest Flows' side, the project aims to create a biophysical model of forest hydrology that accurately predicts water storage and release for entire catchments while also providing data on changes in water quality over time. From the TAIAO's side, the focus lies on developing methods to deal with large volumes of environmental data collected over time and processed incrementally as data streams. Such methods analyse data as it arrives, removing the need to store vast amounts of data for processing at a later date. Therefore, these methods have to address underlying characteristics of data that evolve over time (e.g. due to climatic or ecological changes), and data collected at a range of time intervals and spatial scales ranging from broadscale satellite images to single-point measurements.

So far, we have employed statistical analysis and unsupervised anomaly detection techniques (probabilistic, linear, and proximity-based) to improve data quality. This process was made by cleaning the data from outliers and aimed to achieve highly reliable data currently being used as input for Machine Learning problems. Specifically, the problems of (i) finding similarities and dissimilarities among the planted forests' many logical divisions; (ii) forecasting the time series using complex data combined from multiple sensors.

In future work, we intend to combine the specialist knowledge from SCION's group with the knowledge extracted from the data through Machine Learning and Data Mining processes to model the behaviour of the forests into classes. Once that is done, we intend to combine time-series forecasting with a classification algorithm that will allow the system to correctly model the forests and predict when events like growth spurts, stress, and suppressed trees are bound to happen.

ABOUT THE PRESENTING AUTHOR

Guilherme Weigert Cassales
I am currently a research fellow at the AI Institute from University of Waikato. I am mainly interested in algorithms and tools that employ efficient Machine Learning for Big Data Streams.

Recent Developments and Partnerships in iRODS Technology Enable Global eResearch Collaboration

David Fellinger
iRODS Consortium
davef@renci.org

ABSTRACT

The iRODS (Integrated Rule Oriented Data System) Technology development work can be traced back to a government funded effort in 1995. The initial purpose was to provide management tools for very large data collections. Universities and research institutions began using the technology and that community of users along with storage vendors funded the iRODS Consortium in 2013. Today, the consortium is a professional development organization funded by a vibrant community of over 30 members. The resultant product provides a solution to various data management challenges faced by the diverse community. The primary focus of iRODS is simply to allow researchers to discover data through the use of metadata queries in an expanded collection that may include many disparate file systems. Over a period of time other capabilities have been added that address automatic data ingest and storage tiering but the focus is that all functionalities are metadata driven and automated through rules that describe site policies. Storage facilities and policies may change over time and iRODS can provide the required management to enable the evolution.

Federation capabilities were added as a tool to allow diverse sites to share data. For example, the European Union Data Organization (EUDAT) services 25 research institutions that can share data over the entire EU within legal constraints. CyVerse, which is based in Tucson, has 105,000 users in 160 countries sharing 10 petabytes of storage all under iRODS control.

The demand for collaboration has broadened.

Institutions with geographically diverse sites required that iRODS technology manage replication with parity integrity checking. This provides both disaster recovery capability as well as locality for rapid access. The problem is, of course, that multiple versions of the same file could exist in various locations as researchers modify and add information. The iRODS technology now features worldwide file locking capability. Individual file systems have implemented locks to assure that multiple users cannot inadvertently write to the same file at the same time but iRODS virtualizes and can address many file systems in diverse locations. The new iRODS feature operates so that if a file is opened for "write" anywhere in the world a lock is placed on that file globally. When the "write" operation is concluded iRODS will synchronize all of the replicas worldwide. As a result, a user will never see "stale" data.

Many institutions use "cloud" storage for collaboration purposes so the S3 interface in iRODS has been improved to the point where it is actually faster than native cloud interfaces. This has been tested with all of the major platforms. To further enable data sharing, a partnership with Globus has enabled users to license a plug-in connector developed by the iRODS Consortium. This allows Globus users to manage data with iRODS and share it through a Globus endpoint. Finally, a partnership was launched with OMERO allowing users to share microscopy images that can be discovered through iRODS. In conclusion, iRODS users can now collaborate on a global scale.

ABOUT THE AUTHOR

Dave Fellinger is a Data Management Technologist and Storage Scientist with the iRODS Consortium. He has over three decades of engineering experience including film systems, video processing devices, ASIC design and development, GaAs semiconductor manufacture, RAID and storage systems, and file systems. He attended Carnegie-Mellon University and holds patents in diverse areas of technology.

Common workflows patterns in scientific computing

Alexander Pletzer, Hilary Oliver, Maxime Rio
NeSI/NIWA, NIWA
alexander.pletzer@nesi.org.nz

ABSTRACT

The complexity of scientific computing is increasing. eResearch now often involves a multidisciplinary approach, which may require coupling different software components and/or orchestrating the processing of data from multiple sources. Workflow engines such as Cylc are ideally suited for such tasks. Here, we visit some of the common workflow patterns we have encountered while helping scientists during NeSI consultancies. By far the most common pattern involves the concurrent execution of different models whose performance is summarised in a final task. Another pattern involves running multiple instances of a model in parallel and selecting the results from the model that finishes first, while killing all other instances. We will also discuss how researchers can make their workflows resilient to hardware and other failures, a prime concern for long running simulations.

ABOUT THE AUTHORS

Alexander Pletzer is a High Performance Research Software Engineer working for NeSI at NIWA. Alex helps researchers run better and faster on NeSI platforms.

Hilary Oliver is Principal Scientist – Research Software Engineer at NIWA and the creator of the Cylc workflow engine.

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

Enabling network connectivity with eduroam Visitor Access

Vladimir Mencl

Research and Education Advanced Network New Zealand

vladimir.mencl@reannz.co.nz

ABSTRACT

The eduroam network has been seen as a great success, giving members of the R&E community seamless network connectivity at an ever-growing number of locations. Having to sort out WiFi access at collaboration meetings seems to be a memory from the past – except for the occasional visitor whose organisation is not participating in eduroam yet.

Instead of having to run a parallel guest WiFi network alongside eduroam, visitors who do not have eduroam access can get temporary credentials for accessing eduroam from **eduroam Visitor Access** (eVA), a new service being launched by REANNZ.

The eduroam Visitor Access service is run by REANNZ, on software shared by partner NREN SURFnet. After a REANNZ member organisations (already participating in eduroam) is set up on eVA, authorised staff from this organisation can create credentials for the organisation's visitors. For this, eVA provides an easy-to-use web interface, with authentication handled via Tuakiri, the New Zealand Access Federation.

Credentials can be delivered to visitors in a number of ways, including:

- TXT message
- Email
- Printed handouts for pre-created accounts
- Or the credentials can be requested by the visitor through sending a TXT message with a code to a phone number assigned to eVA

Once configured on the end-user device, the credentials allow the user to connect to the eduroam network in the same way as any other eduroam credentials. Thus, eVA helps bridge the connectivity gap and give all participants equal access to eduroam.

This talk will present the eduroam Visitor Access service in further detail, covering the various use cases eVA supports. Also, the REANNZ team will be at the conference to engage with members interested in adopting this solution.

ABOUT THE AUTHOR

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.

Security vulnerability management in the Nectar research cloud at the University of Auckland

Martin Feller
University of Auckland
m.feller@auckland.ac.nz

ABSTRACT

Responsibility for cybersecurity on the Nectar infrastructure at the University of Auckland is a partnership between researchers and the University.

Vulnerability management is an ongoing process of identifying, assessing, reporting on, managing and remediating cyber vulnerabilities across the research cloud.

Our focus is on managing and mitigating vulnerabilities at scale, raising awareness and increasing literacy in the research community, and helping researchers integrate cybersecurity into their research practice.

To reduce cybersecurity risk and to protect the University and researchers from data breaches and reputational harm we have developed tooling in collaboration with the operational security team. This talk presents the status of the work as well as ongoing challenges.

ABOUT THE AUTHOR

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.

The state of data in research within Aotearoa

Nick Jones, New Zealand eScience Infrastructure,
Michelle Blake, Tahu Kukutai, Rogena Stirling, University of Waikato
Richard Hartshorn, University of Canterbury
nick.jones@nesi.org.nz

ABSTRACT

The Aotearoa New Zealand National Committee on Data in Research (CoDiR) is a newly established group that brings together a diverse range of experts, communities, and organisations involved in the production, measurement, collection, storage, and use of data in research. The Committee provides a unique forum for identifying the challenges, opportunities, and initiatives which characterise Aotearoa's current and future research data ecosystem, and connects to the global research ecosystem through national membership roles within such bodies as CODATA, WDS, and RDA. The committee recognises the uniqueness of Te Tiriti o Waitangi as a guiding and innovative force and supports the principles of Indigenous and Māori data sovereignty.

An initial project has been funded by MBIE to undertake a review of data in research within Aotearoa NZ. This review will:
●Define the area of research data including:
o Defining research data across it's lifecycle
o Impact and value of research data
o Open and Safe data in research
o Indigenous data sovereignty
o Digital research infrastructure
o People, skills, expertise and workforce
● Describe the current national investments and support and look at international benchmarks
● Undertake a national landscape stocktake and present what is operating at a national level identifying research organisations, funding agencies, publishers, research-adjacent organisations as well as third-party service providers and international organisations

The project is at the early stage and this workshop provides an opportunity to hear more about the initial framing for this landscape review and through a World Cafe style approach provides an opportunity to help the project team crowdsource the challenges and opportunities we face as a sector.

What Māori metadata can and should be retained alongside data?

Maui Hudson, Te Kotahi Research Institute, University of Waikato,
Libby Liggins, School of Natural Sciences, Massey University,
maui.hudson@waikato.ac.nz, L.Liggins@massey.ac.nz

ABSTRACT

Metadata is data that provides information about other data. Importantly, for data that is a very reduced representation of a more complex entity (e.g. a DNA sequence of a taonga species, a 2-dimensional map of seamount, a sound recording of many voices in an Indigenous language), metadata provides context. In this way, metadata can enrich the value of the data for those wanting to use the data, those that produced the data, and any others that have an interest or association with the data. Māori have an unparalleled relationship to the physical and natural world of Aotearoa; it includes the metaphysical, familial relationships (whakapapa), and it invokes responsibility for future generations Māori to maintain it. This complete context will never adequately be captured using metadata, but the inclusion of some Indigenous/Māori metadata alongside data can help to avoid complete severance of the data's Indigenous provenance and relevance and help to socialise the importance of Indigenous contexts. Within some existing international metadata standards there are fields that can be co-opted for the inclusion of Indigenous metadata (such as "landowner", "rights"), and there are further initiatives developing new fields (such as "Local Contexts Project ID") where links can be provided to external content where Indigenous communities can provide essential context for the data. However, there are two trade-offs that need to be explored: 1) there is a tension between using already existing fields that are widely used – which will keep Indigenous metadata within existing "core" metadata fields, versus developing fit-for-purpose fields – which may not be picked up by the metadata standards communities; and 2) metadata should ideally be readable/understood by both a human-user and machine-user – how can we cater to both, without being brutally reductionist? In this BoF session we'd like to discuss existing modes for including Indigenous metadata alongside data, how we can support the uptake of Māori metadata in our data and research practice, and what Māori metadata would be ideally retained in future.

ABOUT THE AUTHORS

Maui Hudson (Whakatōhea) is an Associate Professor and Director of the Te Kotahi Research Institute (https://www.waikato.ac.nz/rangahau/) and a co-Director of Local Contexts (https://localcontexts.org/). Maui co-authored the CARE Principles for Indigenous Data Governance, co-developed the BC Labels, and is a founding member of Te Mana Raraunga: Māori Data Sovereignty Network (https://www.temanararaunga.maori.nz/) and the Global Indigenous Data Alliance (https://www.gida-global.org/).

Libby is an Associate Professor at Massey University, and a Research Associate of the Auckland Museum Tāmaki Paenga Hira in Aotearoa New Zealand. She specializes in using genomic data to examine the generation, maintenance, and erosion of marine biodiversity. Libby participates in several initiatives interested in the ethical stewardship and aggregation of data. She is on the Scientific Committee for the International Science Council's World Data System, and the Steering Committee for the Genomic Observatories Metadatabase (GEOME) and the Group on Earth Observations Biodiversity Observation Network (GEOBON) Genetic Composition Working Group.

Institutional approaches to sensitive research data

Yvette Wharton, Laura Armstrong, Mark Gahegan, Waipapa Taumata Rau | University of Auckland
Nick Jones, Claire Rye, Georgina Rae, New Zealand eScience Infrastructure (NeSI)
y.wharton@auckland.ac.nz, l.armstrong@auckland.ac.nz, m.gahegan@auckland.ac.nz,
nick.jones@nesi.org.nz, claire.rye@nesi.org.nz, georgina.rae@nesi.org.nz

ABSTRACT
Sensitive research data can take many forms including personally identifiable and health/medical data, indigenous data, sensitive data about vulnerable species or archaeological sites, and commercial-in-confidence data. It also has different interpretations depending on background and settings, and within and between institutions. These data are commonly subject to legal, ethical and other obligations which impose restrictions on access, use, and handling.

Although most institutions have research involving sensitive data, there are currently no commonly adopted processes, policy or architecture surrounding these data, this leads to difficulty for both researchers and their institutions aiming to support them. The Australian Research Data Commons (ARDC) has recently funded a framework for institutional underpinnings which includes draft recommendations for Sensitive Research Data in Australia (See ARDC - Institutional Underpinnings document, "Draft Framework Element: Sensitive Research Data" https://tinyurl.com/ardc-iu-draft)

This BoF seeks to share experiences with a series of brief introductory remarks followed by a discussion of the challenges and opportunities for collaboration in producing a national, reproducible blueprint that could enable the implementation of sensitive research data infrastructure, strategy, governance and supporting requirements for the FAIR and CARE principles including:
- Researcher perspectives with regards to working with sensitive data
1. research data classification and policy across organisations
2. governance
3. technologies and systems alignment with governance and cybersecurity requirements

ABOUT THE AUTHORS
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. Areas of focus include enabling researchers to manage research data following best practices, including FAIR, CARE and Māori Data Sovereignty data principles, and providing digital research skills and community building. http://orcid.org/0000-0003-2370-3924

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

Mark Gahegan is director of the Centre for eResearch at the University of Auckland, a group that aids researchers to tackle challenging computational research by providing scientific computing and software engineering expertise, patternable services such as virtualised compute platforms, research data services, visualisation and analytics, and educational offerings to up-skill the research community. Mark's research interests are broad, covering e-Science, GIS, visualization, philosophy of science, semantics and pragmatics, geocomputation, machine learning, spatial analysis and spatial data structures. http://orcid.org/0000-0001-7209-8156

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. She has extensive experience in University teaching, research and IT environments and is passionate about using her broad knowledge to facilitate people to achieve their aspirations. http://orcid.org/0000-0002-6689-8840

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 strategic directions and performance overall, bringing together a talented and diverse array of people, and their institutions and interests. https://orcid.org/0000-0001-5513-8312

Building Research Capabilities With The Carpentries

Dr Nisha Ghatak, NeSI
Dr Murray Cadzow, University of Otago
Dr Tom Saunders, CeR, University of Auckland
nisha.ghatak@nesi.org.nz, murray.cadzow@otago.ac.nz, tom.saunders@auckland.ac.nz

ABSTRACT
Computational skills are essential for modern research, but opportunities to learn basic research computing lab skills are few and far between. The Carpentries is a global community of instructors, lesson maintainers, and other supporters who build capacity in data and computational skills for conducting efficient, open, and reproducible research. The Carpentries approach is to build openly-available lessons on research computing, and deliver these lessons using evidence-based teaching practices.

This Birds-of-a feather session will outline the training efforts of the Carpentries community in New Zealand, and explore the future of Carpentries-based training events in Aotearoa. Coming off of an instructor training event and Carpentry Connect, this BoF will encourage discussions on a variety of topics, including how Carpentries workshops can respond to the training needs of the research community.

ABOUT THE AUTHORS
Nisha is the Research Communities Advisor and Training Lead at NeSI. She is also the Carpentries Regional Coordinator for Aotearoa New Zealand.

Murray is a Scientific Programmer on the Research Teaching IT Support team at the University of Otago. Murray has a research background, and now works alongside researchers to help them achieve their computational research goals. In recent years he has been heavily involved in computational literacy and bioinformatic training at the University of Otago - organising ResBaz 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. Murray is a Genomics Aotearoa training associate, and a member of the NeSI Research Reference Group.

Tom is an eResearch Engagement Specialist in the Centre for eResearch, University of Auckland. He has a background in entomology and biological control, but is now working to transfer digital research skills to doctoral students and researchers. He organises and co-teaches Carpentries workshops as a certified instructor, and is involved in organising and teaching in Research Bazaar Aotearoa. Tom maintains an interest in open research, and authored a report on increasing open access rates in Aotearoa during an internship in the Office of the Prime Minister's Chief Science Advisor ("The Future is Open").

eResearch, NeSI and beyond

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

ABSTRACT

Background

NeSI has been a feature on the eResearch landscape for over 4,500 days, or 12.5 years. During that time the sector has remained relatively passive with its eResearch investment, with only one significant institutional eResearch centre emerging. Most institutions remain challenged to make sufficiently substantial investments to realise the benefits touted by eResearch evangelists. As with many aspects of the science system in New Zealand, our fragmentation within the sector compounds our relatively small scale.

Early in 2020 the Government released Te Pae Kahurangi, a review of the Crown Research Institutes. This was then followed in early 2021 by Te Ara Paerangi, a green paper which explored the needs for transformation of the science system. In late 2022 a white paper is being released to map out the future pathways for the sector. Throughout the research sector's discussions and explorations of related issues, there is often a recognition that we have made too little progress in eResearch, and yet it is critical to an ever increasing majority of our science aspirations and needs as a nation.

During 2022 MBIE commissioned the eResearch Report, which took as its subject the exploration of what is needed, to power up our eResearch capabilities nationally. The report explored the future needs of the sector, and the barriers and opportunities needing addressing. It highlighted the successes to date for the current eresearch infrastructures, and the broader impacts being seen through development of researcher skills, shared capabilities, and innovative models of collaboration.

This talk will traverse some of these developments, while aiming to ignite interest in the pathways ahead of us all. It will offer insights from a diverse range of sources on what's shaping our current situation, and present provocations that encourage greater aspiration and action towards meeting our future eResearch needs.

ABOUT THE AUTHOR

Nick Jones Develops innovative teams working with emerging information tech in building advanced capabilities within the research sector. Nick's background is in technology systems, leadership, and governance - designing and building alongside users and partners. He focuses on innovation and technology diffusion, the dynamics of teams and ecosystems, and on leadership and building adaptive organisations.

Taking an Agile Service Management Perspective at NeSI

Georgina Rae, Jun Huh, Matt Chamberlain, Gene Soudlenkov, Nick Jones NeSI
georgina.rae@nesi.org.nz, jun.huh@nesi.org.nz, matt.chamberlain@nesi.org.nz,
gene.soudlenkov@nesi.org.nz, nick.jones@nesi.org.nz

ABSTRACT

Background
Research computing has long been the domain of specialists who often lack a grounding in IT. Research computing platforms are typically built and operated through heroic efforts by researchers whose career path takes them into computing from science. Research computing has become mainstream over the last decade. Stakeholder expectations have shifted towards increased user centricity and increased professionalism of the workforce and services. Meanwhile research computing infrastructure is becoming critical to an increasingly diverse range of users, and is expected to maintain service levels equivalent to any service provider.

Through the course of 2022 NeSI has been co-designing, developing, and rolling out a variety of new services through its new Flexi HPC platform. Flexi HPC as a platform enables a range of partners to marshall investments and expertise, collaborating in meeting researcher needs. With a richer range of services being delivered by a wider group of people, clear processes and performance standards are becoming a necessity.

Improving and standardising service management is a goal for many national initiatives, such as those within the European Open Science Cloud (EOSC), with FitSM providing light weight service management standards to normalise service provision across the EU. EOSC is set to become critical infrastructure for the next phase of scientific research, with similar open science clouds becoming the norm in advanced economies. A full ITSM implementation is both beyond NeSI's resources and likely beyond what is needed by the research sector. Inspired, NeSI's approach is informed by EOSC.

Method/Actions
NeSI applies an Agile approach to our ways of working enabling us to prioritise the work based on user needs in an incremental and pragmatic way. This approach is being used for the implementation of FitSM as well.
We started with a baselining activity to understand where we were at against our initial desired state of maturity. We used this base-lining work, complemented with a pragmatic

view of 'urgent must-haves' from our partners, to build a backlog of tasks which we are now working through in sprints.
It is critical to acknowledge that Service Management is done by and for people and therefore changing our ways of working means being deliberate about navigating through the change as an organisation. This is woven through how we work as a project team.

Results
Despite being at the beginning of the FitSM journey we are already seeing results. A growing focus on having useful, visible, adaptable processes and a useful framework around which to review and draft service agreements.

Conclusions

Agile Service Management is not going to happen overnight, nor is it something that can be done on behalf of the group who will eventually need to implement the processes. By taking an agile perspective we are giving the team the time and customer-driven motivations to genuinely improve how we manage our services.

References https://www.fitsm.eu/

ABOUT THE AUTHORS

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

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

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 strategic directions and performance overall, bringing together a talented and diverse array of people, and their institutions and interests.

Developing eResearch Services - An update on AgResearch and NeSI partnership on building national capability

Nauman Maqbool, Dongwen Luo, AgResearch Ltd
Georgina Rae, NeSI
nauman.maqbool@agresearch.co.nz, dongwen.luo@agresearch.co.nz, georgina.rae@nesi.org.nz

ABSTRACT

AgResearch, in partnership with NeSI, has established an eResearch Platform to address it's researchers' ever increasing needs in data storage, advanced computing as well as skills development, data management and data governance. Over the last twelve months we have been working in developing services to meet the needs of our researchers dealing with complex research questions. This talk will focus on the progress made since the establishment of the eResearch Platform in rolling out its services.

In addition to the computing and infrastructure services, the platform is also involved in supporting researchers with the development of Apps. "Data Integration App" – is one such example for simultaneous analysis of multiple datasets. It is multifaceted from data entry and data wrangling to data visualization, data tabulation and data correlation as well as final report generation. The talk will demonstrate the structure of the App, the main applications with examples, and containerized deployment.

ABOUT THE AUTHORS

Dongwen Luo is Senior Statistician in the Statistics team of wider Digital Agriculture Science group. He has a background in Statistical consultant and Data Science and is passionate about statistical modelling, machine learning and shiny app development.

Nauman Maqbool is Knowledge & eResearch Leader at AgResearch and has been involved in establishing its eResearch Platform. He has a background in Bioinformatics and Data Science and is passionate about eResearch, data management and data governance has He has held various roles in data science and research management.

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

AgResearch's eResearch Infrastructure - a capability development journey with NeSI

Ben Taylor and Blair Bethwaite
AgResearch and NeSI
ben.taylor@agresearch.co.nz, blair.bethwaite@nesi.org.nz

ABSTRACT
In early 2022 AgResearch entered into a partnership with NeSI to help build and support a highly integrated and connected eResearch Infrastructure as an early tenant of NeSI's new Flexible HPC cloud platform. This investment was driven by the need to support AgResearch's ambitious Science Plan and a desire to partner nationally for alignment with recommendations from Te Pae Kahurangi and to foster shared capability build.
At eRNZ'22 we presented on the high-level design of the infrastructure and some of the key services for AgResearch's users. In this presentation we'll provide an update on where we are along the journey. We'll cover some of the juicy technical speeds and feeds, early use-cases, insights into data management and migration, and discuss the challenges we've tackled in building a pathfinder for national delivery of mid-tier HPC (with a tip of the hat to global supply chain issues).

Classifying animal behaviour using machine learning: A collaborative approach to build capability across disciplines

Kevan Cote, Gosia Zobel, AgResearch,
Dale Fletcher, Peter Reutemann, Albert Bifet, University of Waikato AI Institute,
Kevan.cote@agresearch.co.nz, dale.fletcher@waikato.ac.nz, fracpete@waikato.ac.nz,
gosia.zobel@agresearch.co.nz, abifet@waikato.ac.nz

ABSTRACT

The success of New Zealand's agri-sectors is fundamentally linked to animal welfare. Consumers want welfare assurances, however measuring it is not straightforward. Indirect measurement of animal behaviour captured on video is one approach. The application of artificial intelligence (AI) models to video allowed for an exciting collaboration between AgResearch and University of Waikato's AI Institute.

We aimed to use machine learning to automate the identification of specific that an animal might perform. Knowing this project will involve heavy use of computers' graphics processing unit (GPU) for the analysis, we began by setting up a Docker environment. Docker packages programs and models so that they can be shared and run anywhere. We then used Waikato Environment for Knowledge Analysis (WEKA) to attempt a simple classification on a complex video dataset; this was successful in detecting animals but fell short for the behaviour classifications. Next, we shifted focus towards another analysis option: pose detection. This approach allowed us to pick up points on the animal that relate to their behaviour indirectly through movement. While this approach also failed, it presented a valuable learning opportunity; we recognised that the model was not to blame, but rather the video capture approach. Channelling this knowledge, were able to successfully apply the pose detection toolbox to a different video dataset.

AgResearch gained capability in: 1) use of Weka machine learning toolset designed by University of Waikato, 2) application of pose recognition analysis to video datasets, and 3) use of Docker for collaborative modelling. This knowledge was disseminated in an AgResearch-wide workshop. The University of Waikato AI Institute gained valuable information regarding the intricacies and complexities of developing and capturing animal behaviour. Indeed, both institutions gained an understanding of the data requirements (i.e., lighting and video quality, camera angles) to successfully implement pose detection.

The talk will provide a high-level summary of the methodology used, as well as a brief discussion about the learning curves we faced, how these challenges were tackled, and the positive outcomes of the work. Our goal is to ensure that the knowledge and guidelines we developed are a useful resources for others exploring similar collaborations.

ABOUT THE AUTHOR

Kevan Cote has worked for AgResearch since 2019; he is currently with the Bioinformatics and Modelling Team. Having a background as a mechanical design and manufacturing engineer, he saw an alignment between his previous experience with large-scale automated production systems and the large data collection and analysis processes that exist in many scientific applications; this positioned him well to work across multiple teams. He has experience with animal behaviour and personality prediction machine learning models, GPS devices, autonomous drones and 3D printing (anything from skeletal components to electrical enclosures and even grass models!) Some of his most recent research contributions have focused on biodiversity detection using a combination of remote sensing and machine learning. He enjoys bringing his problem-solving knowledge from manufacturing and automation to the digital agricultural research sector.

Detection and Mitigation of Microbiome DNA Presence in Saliva-Derived Whole Genome Sequence Data

BJ Halliday[1], D Markie[2], E Goodin[1], M Black[3], A Sporle[4], H Watson[5], P Wilcox[6], SP Robertson[1]
Department of Women's and Children's Health, Otago Medical School, Division of Health Sciences, University of Otago, Dunedin, NZ.
Department of Pathology, Otago Medical School, Division of Health Sciences, University of Otago, Dunedin, NZ., Department of Biochemistry, School of Biomedical Sciences, Division of Sciences, University of Otago, Dunedin, NZ., Department of Statistics, University of Auckland, Auckland, NZ.
Ngāti Porou Hauora, Te Puia Springs, NZ., Department of Mathematics and Statistics, Division of Sciences, University of Otago, Dunedin, NZ., benjamin.halliday@otago.ac.nz

ABSTRACT

DNA extracted from saliva is suitable for high-throughput sequencing studies and is crucial in situations where taking a blood sample is not practical or appropriate. However, using saliva samples risks oral microbiome DNA contamination, which may compromise downstream applications of sequencing data. As part of the Genomics Aotearoa funded Variome project, a subset of anonymised whole-genome sequencing data derived from either participants' saliva (n=198) or blood (n=9) samples were assessed. The Variome project aims to assist genomic diagnostics in health care in Aotearoa by generating a population specific genomic catalogue to more accurately represent allele diversity.

Preliminary cohort screening revealed oral microbiome presence in saliva samples, with these sequences estimated to consist of between 4-56% non-human DNA. Misalignment of bacterial sequence to the human reference genome was seen in affected samples, observed to induce false positive variant calls and other quality anomalies. The presence of this bacterial derived variation presents a considerable risk to the goals of the Variome project, as this variation appears novel when compared to other human allele frequency databases and occurs consistently across affected samples.

To ensure a robust and representative Variome resource, a mitigation strategy was developed using the taxonomic classification tool Kraken2 to partition sequence data, allowing for the quarantine of non-human sequences. Consistent with expectation, data derived from blood samples were found to have greater than 99% of sequences classified as human. For saliva samples, non-human sequences were found to originate primarily from oral microbiome bacteria, showing taxonomic consistency between affected samples. However, approximately 20% of non-human sequences in saliva samples were unclassified using the standard Kraken2 classification database.

Novel human sequence may also be present in the unclassified read fraction, so removal of these sequences is inappropriate as it may impact downstream analyses such as novel structural variant resolution. Consequently, unclassified sequences were assembled into contigs using the metagenome assembler MEGAHIT, and subsequently compared to the NCBI nucleotide collection using BLAST+. Sequences from identified taxa were collated and then used to construct a custom Kraken2 database. Previously unclassified taxa primarily consisted of bacteriophage and food sources not represented in the standard Kraken2 database.

Reclassification of sequences using a custom Kraken2 database and subsequent alignment of human and unclassified reads led to a significant decrease in erroneous misalignment. This method is theoretically portable to any target species and can be used to mitigate any non-target species contamination. However, the additional computation required for custom database construction and sequence classification prior to alignment may be impractical for some sequencing studies.

Pangenome graph: towards a "reference freer future"

Zoe Yang, Una Ren, Joep de Ligt, ESR, Kenepuru, NZ
Pjotr Prins, Erik Garrison
Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Tennessee, USA
Mik Black, Department of Biochemistry, University of Otago, Dunedin, NZ.
zoe.yang@esr.cri.nz

ABSTRACT

Whole genome sequencing (WGS) has revolutionized infectious disease surveillance: its utility is especially evident in the SARs-CoV-2 pandemic [1]. Currently, genomic surveillance concentrates primarily on monitoring genomic lineages and establishing linkages between cases. Analysis is dependent on the core genome, that is, regions that are common to all samples being analysed, defined via a single "reference genome" [2,3]. As bacterial genomes are flexible, genomic rearrangements and large-scale deletion or insertion events are common [4], variations in the accessory genome that are only existed in part of the strains or isolates are not typically detected. Understanding the accessory genome and associated variations is important because some accessory genes confer an adaptive advantage in changing environments [5], while others are responsible for pathogenicity and antibiotic resistance [6,7]. Bacterial pangenome analysis has key practical applications, supporting the selection of antimicrobial targets and vaccine candidates [8,9]. Pangenome Graphs [10,11] provide a mechanism for more holistic and unbiased analysis of genomic variation. Moving from using a single genome to a pangenome improves variant calling and identification of genes associated with key traits [12,13]. We will present a bioinformatic pipeline for pangenome graph construction and variant calling, using both simulated data and data from Neisseria meningitidis to demonstrate its function. These results will be used as the basis to discuss potential application to a broad range of microbial pathogens.

ABOUT THE AUTHOR

Zoe Yang- I obtained my Ph.D in Botany from the Institute of Botany, Chinese academy of Sciences. Before I moved to NZ, I spent five years on Cancer Genomics and Evolution at Beijing Institute of Genomics, CAS. Now I am working as a senior scientist at the Institute of Environmental Science and Research (ESR), focusing on population genomics and evolution of bacterial pathogens based on evolution theories, multi-Omics sequencing and bioinformatic analysis. I am interested in tracing evolutionally history and population dynamics of pathogens, identifying genetic variations and phenotype variations and their relations, providing insights and evidence for disease control and treatment. One of my current works is developing the pangenome graph pipeline for Pan Genome construction and variant calling with a focus application on microbial pathogens.

REFRENCES

1.Geoghegan JL, Ren X, Storey M, et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. Nat Commun. Dec 11 2020;11(1):6351.
2. Retchless AC, Fox LM, Maiden MCJ, et al. Toward a Global Genomic Epidemiology of Meningococcal Disease. J Infect Dis. Oct 31 2019;220(220 Suppl 4):S266-S273.
3. Rodgers E, Bentley SD, Borrow R, et al. The global meningitis genome partnership. J Infect. Oct 2020;81(4):510-520.
4. Darmon E, Leach DR. Bacterial genome instability. Microbiol Mol Biol Rev. Mar 2014;78(1):1-39.
5. Vernikos G, Medini D, Riley DR, Tettelin H. Ten years of pan-genome analyses. Curr Opin

Microbiol. Feb 2015;23:148-154.

6. Rasko DA, Rosovitz MJ, Myers GS, et al. The pangenome structure of Escherichia coli: comparative genomic analysis of E. coli commensal and pathogenic isolates. J Bacteriol. Oct 2008;190(20):6881-6893.

7. Holt KE, Wertheim H, Zadoks RN, et al. Genomic analysis of diversity, population structure, virulence, and antimicrobial resistance in Klebsiella pneumoniae, an urgent threat to public health. Proc Natl Acad Sci U S A. Jul 7 2015;112(27):E3574-3581.

8. Tettelin H, Masignani V, Cieslewicz MJ, et al. Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A. Sep 27 2005;102(39):13950-13955.

9. Argante L, Tizzoni M, Medini D. Fast and accurate dynamic estimation of field effectiveness of meningococcal vaccines. BMC Med. Jun 30 2016;14:98.

10.Paten B, Novak AM, Eizenga JM, Garrison E. Genome graphs and the evolution of genome inference. Genome Res. May 2017;27(5):665-676.

11.Chen S, Krusche P, Dolzhenko E, et al. Paragraph: a graph-based structural variant genotyper for short-read sequence data. Genome Biol. Dec 19 2019;20(1):291.

12.Colquhoun RM, Hall MB, Lima L, et al. Nucleotide-resolution bacterial pan-genomics with reference graphs. bioRxiv preprint doi: https://doi.org/10.1101/2020.11.12.380378. 2021.

13.Letcher B, Hunt M, Iqbal Z. Enabiling multiscale variation analysis with genome graphs. bioRxiv preprint doi: https://doi.org/10.1101/2021.02.03.429603. 2021.uffrein-Roberts, S., Allelic expression patterns in psychatric candidate genes. PhD Thesis in Pathology. 2008, University of Otago: Christchurch. p. 216

War on Wildings: continuing Flow Transport Model development

Chris Scott, NeSI
Jiawei Zhang, Scion
chris.scott@nesi.org.nz, Jiawei.Zhang@scionresearch.com

ABSTRACT
In 30 years, approximately 7.5M hectares of land in New Zealand is estimated to be at the risk of invasion by wilding conifers. Successful wilding control can provide up to $6.3B of economic and environmental benefit. The Flow Transport Model (FTM), an analytical model developed by Scion and their CSIRO collaborators, enables seed dispersal modelling which includes the impact of the atmospheric turbulence caused by the complex terrain and vegetation canopy. The model can be used to identify the spatial extent and distribution of the long-distance conifer seed dispersal from any chosen location. Such information can be used operationally to provide information for wilding conifer management in the fight against the invasive species..

During this presentation we will share our experience as we teamed together and continued the technical development of the FTM. We have tried to make the model more accessible and practical for operational use, including adding a simple GUI, while retaining the ability to run via the command line for when HPC resources are required. We have also added a generic input module that simplifies modelling different regions or wind directions as well as different seeding locations. Bottlenecks in performance were identified using profiling tools and we used a combination of Numba and vectorisation to gain significant performance improvements. Cross platform support was important and we have setup a continuous integration pipeline to ensure tests are run on different operating systems. As a result of these improvements the FTM is now more accessible and easy to use and has a much faster time to solution.

ABOUT THE AUTHORS
Chris Scott is a NeSI Research Software Engineer based at The University of Auckland

Jiawei Zhang is an Atmospheric Scientist in Scion based in Christchurch

BoF: Building communities of practice

Jorge Bornemann, Richard Dean, Natalie Forsdick, Alvaro Orsi, Aleksandra Pawlik, Maxime Rio, Katie Rosemond, Annie West, Noel Zeng

ABSTRACT / INTRODUCTION
Communities of practice (COPs) are extremely useful to share knowledge, to seek advice, to instigate new ideas and also to create a sense of kotahitanga/togetherness between practitioners. This BoF session will explore ideas on how to help develop resilient COPs, address the common challenges and promote inter-institutional COPs, amongst other topics.
This session will be of interest for those organising, or intending to organise, self-help groups in their institutions or across several institutions, to those that have an interest in community education and to those with a passion for training and people development.
In the session, each of the panellists will do a short presentation, covering 3 areas: a description of the COP they manage and its scope, what has worked well, and what challenges are they facing. This information will then be discussed in breakout groups with different foci and the key challenges and solutions collated afterwards.

ABOUT THE AUTHOR(S)
Jorge Bornemann: Research Software Engineer, National Institute for Water and Atmospheric Research (NIWA)

Richard Dean: Raraunga Kaipūtaiao / Senior Data Scientist, Institute of Environmental Science and Research Limited (ESR)

Natalie Forsdick: Postdoctoral Research Fellow at Manaaki Whenua – Landcare Research and co-chair of the Genomics Aotearoa Postdocs Group

Alvaro Orsi: Data Science Lead, Institute of Environmental Science and Research Limited (ESR)

Aleksandra Pawlick: eResearch Capability Specialist at Manaaki Whenua – Landcare Research

Maxime Rio: Data Scientist, National Institute for Water and Atmospheric Research (NIWA) and Data Science Engineer, New Zealand eScience Infrastructure (NeSI)

Katie Rosemond: Data Analyst – SME (data), ROBUST, Data & Digital Group, Bureau of Meteorology, Australia

Annie West: Postdoctoral Research Fellow at Waipapa Taumata Rau – University of Auckland, NZ Microbiology Society Executive Committee member and co-chair of the Genomics Aotearoa Postdocs Group.

Noel Zeng: Noel Zeng, eResearch Solutions Specialist, Centre for eResearch, Waipapa Taumata Rau, University of Auckland

Learn how to accelerate your research with Machine Learning on AWS

Shivonne Londt (AWS), Karl de Borst (AWS)

ABSTRACT
Take part in a hands on workshop that will introduce you to AWS machine learning services, followed by a deeper dive into using AWS for common machine learning tasks. Amazon SageMaker is the machine learning service on AWS that can help you to build, train, and deploy machine learning models for any use case with fully managed infrastructure, tools and workflows. Bring your own laptop with a working internet connection and the ability to connect to external websites. We will provide the AWS accounts. No prior machine learning or AWS experience needed - bring an open mind. The workshop will show you how to use a low-code, visual interface to generate predictions from a dataset using Amazon SageMaker Studio. We will move onto using Amazon SageMaker as the development environment for machine learning on AWS and work through some common machine learning tasks, such as preparing, training and deploying machine learning models.

Advancing Human Genomics data sharing in Australia: Highlights from Australian BioCommons' Human Genome Informatics initiative

Marion Shadbolt, Jessica Holliday, Uwe Winter, Steven Manos, Jeff Christiansen, Andrew Lonie, Bernard Pope, Australian BioCommons
marion@biocommons.org.au

ABSTRACT

Background/Context

Currently, the management and sharing of human genomics data in Australia is fragmented and largely siloed within the major national institutes that generate human genomic sequence data. The Australian BioCommons' Human Genome Informatics (HGI) initiative aims to address this challenge through collaborative projects and initiatives with research institutes, infrastructure partners, and government agencies. Here we describe two of the major projects that we are currently undertaking to enhance Findable, Accessible, Interoperable, Reuseable (FAIR) human genomics data sharing in Australia.

The Human Genomes Platform Project

The Human Genomes Platform Project (HGPP) aims to leverage international best practice technologies and global standards to accelerate FAIR human genomics data sharing in Australia. Involving leading Australian human genomics research organisations, along with national computing infrastructure partners, the HGPP is facilitating the deployment of much needed genomic data sharing infrastructure in Australia. Specifically, the project is investigating which existing international solutions meet Australian requirements, with sub-projects aligning to GA4GH and other internationally developed software and standards (Beacon, CILogon, Resource Entitlement Management Software, Federated EGA). The HGI team provides strategic leadership, project coordination, technical expertise and domain knowledge to this national, multi-institutional and collaborative platform project.

The Australian Cardiovascular disease Data Commons

The Australian Cardiovascular disease Data Commons (ACDC) will be a comprehensive, secure, scalable, internationally integrated data infrastructure connected to global best practice analysis platforms, to enable the identification of novel insights and predictive biomarkers for Coronary Artery Disease (CAD). Currently in a pilot phase, the Australian BioCommons has provided coordination, technical and domain knowledge expertise to progress the project. The HGII and Cyberinfrastructure teams have been instrumental in establishing a pilot data portal powered by Gen3 and comprising a custom data dictionary, as well as developing custom tools and pipelines to facilitate its use. We have also facilitated Gen3 platform development and integration with key features such as federated identity and access management through CILogon and COManage.

ABOUT THE AUTHORS
Marion Shadbolt
Since completing her Masters of Bioinformatics at the University of Melbourne in 2016, Marion's career has spanned three continents and a range of technical and informatics support roles.

At the Karsan lab at BC Cancer (Vancouver, Canada), Marion developed knowledge and expertise in clinical genomic assay development and analysed various sequencing datasets from blood cancer patient samples. During this experience, Marion encountered frustrating inconsistencies in publicly accessible datasets and gaps in the understanding of FAIR data practices at the research lab level.
In 2019, she joined EMBL-EBI (Hinxton, UK) to work on the Human Cell Atlas (HCA) project to build her knowledge and experience in this field. As part of the Data Coordination Platform team, she worked with collaborators and research labs to facilitate harmonisation and ingestion of vast quantities of cellular resolution data and metadata.

Returning to Australia in 2021, Marion took up the position of Human Genomics Data Specialist at the Australian BioCommons. Within the Human Genome Informatics team, she collaborates with national partners and international colleagues to facilitate human genomic data sharing in Australia.

Data Stewardship Wizard: A tool for Digital Data Management Plans

Libby Li[1]

[1]Waipapa Taumata Rau | University of Auckland,Tāmaki Makaurau Auckland, Aotearoa New Zealand

libby.li@auckland.ac.nz

ABSTRACT

It is now more common, both nationally and internationally, for funders and ethics committees to require researchers to provide a data management plan (DMP) as part of their application process. We have looked at some of the DMP tools available that support researchers developing data management and sharing plans, such as Data Stewardship Wizard (DSW), DMPonline, and DMPTool.

Among these tools, DSW is outstanding, providing a simple question-based way to create a DMP - researchers don't have to write a lot of text. It also enables collaborative development of machine-actionable DMPs with tools that allow questions, comments and co-writing.

This talk will discuss our experience setting up a self-managed instance of the DSW tool for a genomics research group, and their feedback using this tool. DSW as a digital tool for data management planning is promising and can be used not only in genomics but also other domains.

ABOUT THE AUTHOR

Libby Li is an eResearch Solutions Specialist at Waipapa Taumata Rau | University of Auckland with a background in bioinformatics

Supporting a national data community - the Committee on Data in Research (CoDiR)

Maryam Alavi, Plant and Food Research Michelle Blake, University of Waikato Bapon Fakhruddin, Tonkin + Taylor
Richard Hartshorn, University of Canterbury Nick Jones, NeSI
Tahu Kukutai, University of Waikato Andrew Lensen, Victoria University of Wellington
Barry Milne, University of Auckland Gerald (Jay) Naepi, Matada Research Group
Grace Walker, University of Otago

ABSTRACT

Background/Context

In late 2021, following New Zealand becoming a member of CODATA, the Royal Society Te Apārangi, as the member body, sought a mechanism to appoint its delegate to CODATA (and a delegate to the WDS, of which NZ is also a member). As there was no national body that covers research data in its broad sense, the Royal Society Te Apārangi set about formation of a National Committee. Approximately a dozen data-related organisations were invited to identify candidates to explore its formation. Discussion at the wider meetings indicated that the group anticipated such a body would undertake a broader role in connecting and facilitating data in research, including it's production, measurement, collection, storage, and use.

In early 2022 an establishment group of six people from a range of backgrounds and disciplines was identified, while a broader group that would come to be known as the Forum maintained an active interest in ongoing developments. The establishment group were asked to outline how such a committee might operate and sketch out terms of reference.

During their formative meetings, the possibility of the committee initiating projects was discussed and a number of possibilities were identified. One of these (a data landscape scoping paper) was clearly delineated and appeared central to much further work in the space. The idea was pitched to the Ministry of Business, Innovation and Employment (MBIE) and received a small amount of funding, with members of the establishment group as the steering group.

The proposed terms of reference for the Aotearoa New Zealand Committee on Data in Research (CoDiR) were presented to and approved by consensus at a meeting of the wider group (the CoDiR Forum). Expressions of Interest were then sought from people interested in joining CoDiR, with explicit statements that we were looking for diversity of discipline,

organisation, background, gender, and career stage. Approximately 30 candidates applied, including around 10 early career researchers. The founding membership for CoDiR was presented to and endorsed by the wider forum in mid-2022 (two ECRs were selected).

Related work

CoDiR provides a point of coordination between national and international practitioners - experts, communities, and organisations involved in the production, measurement, collection, storage, and use of data in research. Internationally, it connects with CODATA, World Data System, and Research Data Alliance.

Method

CoDiR connects and promotes national and locally practitioners with a wide range of interests in Data in Research, including:

- promotes principles, policies and practices for Open Data, FAIR data, Open Science,

and CARE principles for Indigenous data governance, particularly within the context of Te Tiriti o Waitangi, Māori data sovereignty, and collective well-being;

- links Aotearoa New Zealand to global practices and advances in data quality, measurement, management, analysis, curation and preservation, through cultivating relationships with and membership of appropriate international organisations and projects;
- ensures that Aotearoa New Zealand has a voice and a role in the global development of the ethical use of data in research, and Indigenous data sovereignty;
- promotes the adoption of best practices by researchers working with data in Aotearoa New Zealand;
- organises and promotes activities such as workshops and panels, to promote national and international collaborations and to raise awareness of policies, technologies, infrastructure and best practices in data in Aotearoa New Zealand;
- contributes to the development of strategic policy advice on the production, collection, storage, dissemination, sharing, and use of data in research;

Results

CoDiR is in an establishment phase. It has an endorsed terms of reference, a membership from a wide range of origins and experiences, a Forum and its first couple of projects under development. Alongside highlighting the role of the Committee and the Forum, the work underway on the data in research landscape will be discussed, along with other emerging themes and interests the committee is involved in.

REFRENCES
CoDiR Terms of Reference https://dir-aotearoa.github.io/committee/

ABOUT THE AUTHORS

Richard Hartshorn
An international expert in the field of chemical nomenclature, a consultant for chemistry educators and textbook authors around the world. Passionate about science outreach. Research interests include anti-cancer drug design and systems of photoactivated delivery.

Maryam Alavi
Data scientist, passionate about using analytics to develop data-driven recommendation-making tools to support better decision-making in complex issues.
Experience in identifying practical problems in industry and public sectors, finding solutions and communicating with wide range of stakeholders.

Michelle Blake
Firm believer in developing and empowering staff to reach their full potential. A skilled relationship builder, who acts as a stimulator and connector of people and ideas to create high impact outcomes. A champion of user experience techniques, engaging with communities to implement co-created outcomes.

Bapon Fakhruddin
An international climate change and disaster risk reduction expert. Experience with climate-vulnerable countries around the world, including in the South Pacific and New Zealand. Expert in climate resilience and responsiveness at the community, government and inter-agency level.

Nick Jones
Develops innovative teams working with emerging information tech in building advanced capabilities within the research sector. Nick's background is in technology systems, leadership, and governance - designing and building alongside users and partners. He focuses on innovation

and technology diffusion, the dynamics of teams and ecosystems, and on leadership and building adaptive organisations.

Tahu Kukutai
Researcher on Māori demography and Indigenous data sovereignty, working with hapū and iwi, doing demographic research that meets their needs. A founding member of the Māori Data Sovereignty network Te Mana Raraunga and the Global Indigenous Data Alliance.
Serves on a wide range of iwi, research and government advisory groups including the Pūhoro Charitable Trust which governs the Pūhoro STEMM Academy, the Chief Science Advisor Forum, and the technical advisory for the Data Iwi Leadership Group, National Iwi Chairs Forum.

Andrew Lensen
He works in Artificial Intelligence centred around unsupervised learning, which is all about discovering underlying patterns in data — and what use is such a discovery if we cannot easily understand it. He is also interested in the application of AI within an Aotearoa context, across governmental and industrial applications on tasks that will affect our everyday lives including the use of AI in biology, legal decisionmaking, psychology, and more.

Barry Milne
Barry's research focus is on life-course and longitudinal studies, and using large or complex whole population datasets to answer a broad range of policy and research questions – from analysing the influence of family medical history on the development of psychiatric disorders to looking at what childhood factors can be intervened upon to ensure positive developmental outcomes.

Gerald (Jay) Naepi
Experience in healthcare, insurance, fitness, and wellness, leading to a career of helping others. Through Matada Research, applies business strategies that adhere to values and objectives with the ultimate goal of improving Pacific health and wellbeing.

Grace Walker
A Maori data scientist who has a background working with longitudinal data sets. Engages in projects focused on supporting the uptake of data within communities and using data to change the narratives of indigeous and minority populations. Lead of the Maori housing framework with Te Matapihi. A strong advocate for supporting mahi that positively utilises data to benefit communities to enhance mana motuhake as well as encourage the uptake of STEM amongst indigenous and underrepresented groups.

FAIR Capabilities: Quantifying the benefits of institutional repositories for graduate students' research outputs

Andrew Mckenna-Foster
Figshare
andrew@figshare.com (Please email laura@figshare.com with notification of acceptance or rejection).

ABSTRACT

Electronic theses and dissertations (ETDs) are important additions within institutional repositories around the globe. The electronic versions of theses or dissertations are more findable and accessible to end users thus providing benefits, like higher visibility, citation, and reuse, to both the author and the institution. The relatively new paradigm of open data, and the technologies and skills that enable it, add a new dimension to ETDs. Openly shared data may increase the reuse of an ETD and increase the credibility of the conclusions by easing reproducibility.

As potential early career researchers, graduate students should take advantage of these benefits by developing capabilities in open data skills and familiarity with the FAIR principles. Their success may be influenced by research data management (RDM) support and open data policies from their institutions, funders, and publishers. They may be sharing outputs through an institutional repository, potentially with help from curators, or directly in a generalist repository.

Objective: Our question was "Can we see an impact on the FAIRness of records shared through an institutional repository compared to a generalist repository and how do universities in New Zealand and Australia compare?" Specifically, we wanted to compare the FAIRness of ETD related outputs from universities in New Zealand and Australia to a global sample of ETD records presumably shared directly by students, in a generalist repository, as well as to those shared in other institutional repositories.

Methods: We programmatically collected ETD related records from the Figshare platform universe of repositories. Our study uses proxy measures for metadata richness and interoperability to assess FAIRness.

Results: Preliminary results suggest that New Zealand and Australian repository records do have richer metadata than records shared by individual students and relatively similar metadata quality to records from other universities. Records from institutional repositories have significantly higher views, suggesting higher reuse.

Conclusions: Our presentation will explore these results in the context of RDM training and the future of data reuse. We also suggest that our methods can be used to assess how data sharing capabilities in students change over time as knowledge of the FAIR principles expand and data sharing becomes standard practice.

Preliminary results:
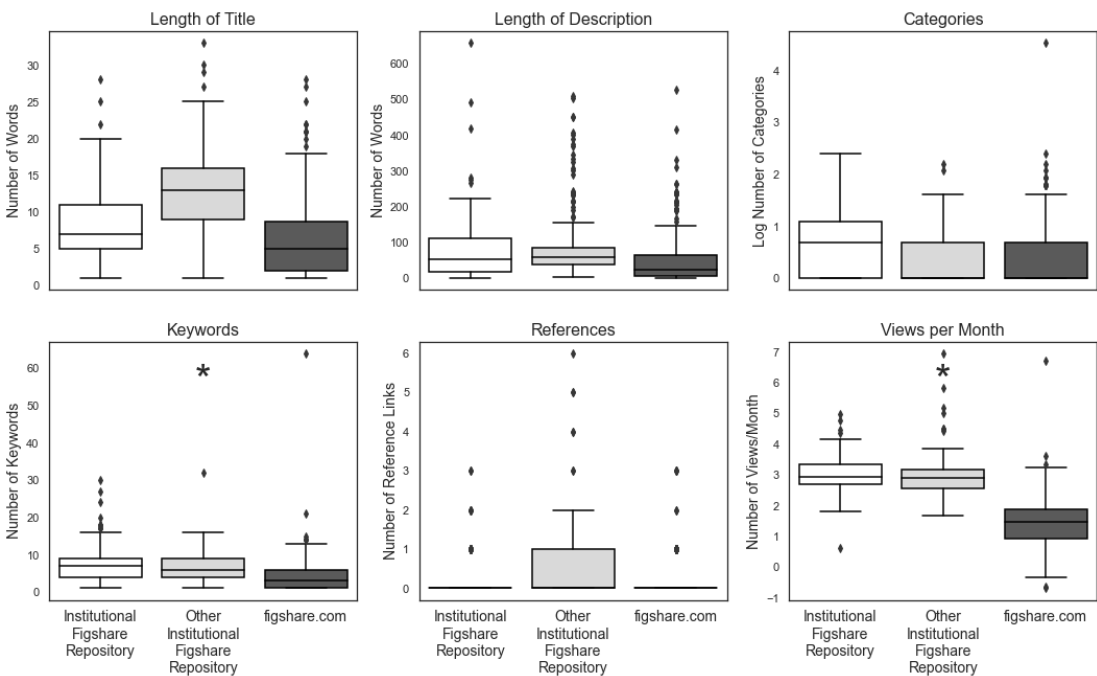
Here, "Institutional Figshare Repository" refers to repositories in Australia or New Zealand. There are some significant differences.

ABOUT THE AUTHOR

Andrew Mckenna-Foster is a Product Specialist at Figshare with over 12 years of experience working with research and museum data. For a decade, Andrew oversaw the operations of a small natural history museum and aquarium and directed the related research programs and biological collections. During this time, he published research on spider ecology and endangered species conservation. An interest in open science and data management led Andrew to the field of information science and he received a masters of library and information science, with a focus in data curation, from the University of Washington in 2020. At Figshare, Andrew focuses on advancing the practice of open science and looks for ways to help researchers and librarians think through their open research and repository needs

How will the GPU in your pocket change the future of eResearch?

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

ABSTRACT
Deploying machine learning models trained on GPUs typically requires access to high performance compute. However - thanks to libraries such as tensorflow lite, tensorflow.js and google's MLKit, we can deploy models to edge devices and mobile phones.
How will the GPU in your pocket change the future of eResearch?
In this presentation, Richard will explain how the COVID-19 pandemic provided an opportunity for ESR to deploy machine learning models to mobile GPUs, and how the result democratises access to cutting edge science well beyond our infectious disease use case.
Richard will talk about developing the *capability*, how that improves data quality and *connectivity* by removing the need to transport samples to a physical lab and how *equity* is improved by putting cutting edge testing in the hands of the masses.

ABOUT THE AUTHOR
Richard Dean is a senior data scientist at ESR. He works across the organisation on projects that gain insight from big data sets to tackle nitty gritty real-world problems affecting human communities in Aotearoa, covering everything from forensic science to human health and the environment. He has a BSc in Information Systems Management from Durham University and wrote an MSc thesis on public health data interoperability standards while working at the Wolfson Research Institute for Health and Wellbeing in Durham.
Richard was the first member of staff at Public Health England to graduate from the UK government digital service 'data science accelerator' programme. In 2019, he brought the scheme to New Zealand through an internal accelerator programme within ESR which has since trained three cohorts of data scientists.

Hyperparameter optimisation for Machine Learning:
A little bit of theory and practice

Maxime Rio[1,2], Alan Tan[2]
1NeSI, 2NIWA
maxime.rio@nesi.org.nz, alan.tan@niwa.co.nz

ABSTRACT
Fitting a machine learning model can require substantial adjustments to parameters used in order to obtain optimal results. This model tuning phase, also called hyperparameter optimisation, is a very time consuming phase that can - and should - be automated.
In this talk, we will introduce the common techniques, with a focus on applicability, for hyperparameter optimisation. We will discuss each technique's strengths and weaknesses, which depends on the type of model to tune as well as the computing platform available. This includes the potential for parallelisation, to leverage a massively parallel infrastructure such as an HPC, as well as the efficiency for a given compute budget. Finally, we will provide an illustration in a context of models developed at NIWA to improve our weather station related products.

ABOUT THE AUTHORS
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.

Alan is a data scientist with the Artificial Intelligence and Deep Learning team in NIWA. His interest is in applying machine learning for developing solutions that can help researchers automate and scale their work.

Using HPC-driven climate simulations and open-source software to predict future tropical cyclones

Dr Jonny Williams, Dr Erik Behrens, Dr Olaf Morgenstern
NIWA, Wellington
jonny.williams@niwa.co.nz, erik.behrens@niwa.co.nz, olaf.morgenstern@niwa.co.nz

ABSTRACT

Tropical cyclones in the South Pacific are few (~< 10 per year [NIWA, 2022]) but can cause catastrophic damage to property and infrastructure as well as leading to significant loss of life. In this work we present storm-tracking analysis using an open-source Python tool [Oliver, 2018] adapted from a similar ocean eddy tracking code and apply it to observations [Hersbach et al., 2020] and to contemporary climate and future projections [Sellar et al., 2019; Behrens et al., 2020]. The tracking code itself is elegant and simple, requiring only 1 input field (mean-sea-level-pressure), and runs fast without the need for multiprocessing or additional pre-processing. Our results are in broad agreement with previous studies which show that tropical cyclones affecting New Zealand are likely to decrease in number by the end of the century but are predicted to increase in severity. Indeed, depending on the magnitude of end-of-century warming, we predict an increase in energy dissipation per storm of between ≈ 10 - 40%.

References
NIWA, Southwest Pacific Tropical Cyclone Outlook - October 2022, https://niwa.co.nz/climate/southwest-pacific-tropical-cyclone-outlook/southwest-pacific-tropical-cyclone-outlook-october-2022

E. C. J. Oliver, https://github.com/ecjoliver/stormTracking, 2018.

H. Hersbach et al., The ERA5 global reanalysis., Q. J. RMS., 2020.

A. Sellar et al., UKESM1: Description and Evaluation of the U.K. Earth System Model, JAMES, 2019.

Behrens et al., Local Grid Refinement in New Zealand's Earth System Model: Tasman Sea Ocean Circulation Improvements and Super-Gyre Circulation Implications, JAMES, 2020.

ABOUT THE AUTHORS

Jonny Williams is a climate scientist at NIWA working with global climate models. Jonny is a strong advocate of reproducible and open-access research as well as creative data visualisation. Erik Behrens is a physical oceanographer and ocean modeller, with expertise in high-resolution ocean modelling, Lagrangian particle tracking and climate extremes.

Olaf Morgenstern is an expert in global climate modelling, with interests spanning climate and atmospheric science. He is a Lead Author of the most recent 6th Assessment Report of IPCC.

The dynamics of Dictystelium discoideum's microbiome and genomic changes of the symbiont Paraburkholderia agricolaris

Matthew Adlam, Elizabeth Ostrowski
Massey University Albany
M.Adlam@me.com, E.Ostrowski@massey.ac.nz

ABSTRACT

Many Iwi has begun restoring pastoral lands to wetlands and restoration is a key factor in restoring the Mauri of these lands. The microbiome is known to be a key factor in these restorations and taxonomic composition can act as a proxy for wetland health (Birnbaum and Trevathan-Tackett , Gao, Liu et al. 2021). Moreover, these analyses provide a more holistic view of the Whenua which is more aligned with matauranga Maori and the Maori worldview (Parsons and Fisher 2020, Gao, Liu et al. 2021). It is my opinion that Tangata Whenua is directly involved in these studies and has ownership of them. Such studies are becoming more and more affordable. This provides an ideal opportunity for Iwi to get more directly involved in the wetland restoration. The methods I have used in my studies can be applied to monitor and then aid in directing wetland microbiome composition to help generate a healthy and sustainable wetland. The goals of my studies are to describe the diversity and dynamics of the microbiome associated with Dictyostelium discoideum. To address this question, I am examining the microbiome of the social amoeba, a soil-borne amoeba that can exist in either a single-celled state or as a multicellular organism. Prior work has shown that natural isolates of D. discoideum commonly harbour several different bacterial species (Brock, Haselkorn et al. 2018, Haselkorn, Jimenez et al. 2021). Of these, three species of Paraburkholderia have been consistently identified in approximately 1/3 of wild D.discoideum isolates (Brock, Read et al. 2013, Brock, Noh et al. 2020). Two of these species, P.hayleyella and P.bonniea, exhibited evidence of genome reduction. Noh et.al. proposed that these two species have become stable symbionts (Noh, Capodanno et al. 2022). The third species, P.agricolaris, did not show any significant genome reduction. These bacterial isolates were from only one site. However, another study revealed multiple haplotypes of P.agricolaris over a broad geographic range (Haselkorn, DiSalvo et al. 2019).

Here, I describe a large metagenomic analysis of the D. discoideum microbiome. The study is based on 672 natural isolates of D. discoideum that were collected over a large geographic scale ranging from metres to more than 1000 km, with repeated sampling of some sites over three years. I expect that some species may have a stochastic association, whereas others may be stable. I am also investigating differences in the genomes of the amoeba-associated P. agricolaris from geographically distant sites. This study is potentially of great interest as it may show different processes involved in genome reduction as symbiotic associations become stronger and more intimate (Moran 2003, Husnik and Keeling 2019, Noh, Capodanno et al. 2022).

ABOUT THE AUTHORS

Dr. Matthew Adlam, Completing a MSc at Massey Albany. Prior work in molecular Immunology PhD (2001) and Immunology until 2009 . I am now re-training as I find the bacterial world of metagenomics fascinating. I am hoping to combine my immunology skills and knowledge with the fascinating emerging field of microbiome research.

Dr. Elizabeth Ostrowski is a Senior Lecturer in Ecology and Zoology at Massey University in Albany.

REFRENCES

Birnbaum, C. and S. M. Trevathan-Tackett "Aiding coastal wetland restoration via the belowground soil microbiome: an overview." Restoration Ecology: e13824.

Brock, D. A., T. S. Haselkorn, J. R. Garcia, U. Bashir, T. E. Douglas, J. Galloway, F. Brodie, D. C. Queller and J. E. Strassmann (2018). "Diversity of Free-Living Environmental Bacteria and Their Interactions With a Bactivorous Amoeba." Frontiers in Cellular and Infection Microbiology 8(411).

Brock, D. A., S. Noh, A. N. M. Hubert, T. S. Haselkorn, S. DiSalvo, M. K. Suess, A. S. Bradley, M. Tavakoli-Nezhad, K. S. Geist, D. C. Queller and J. E. Strassmann (2020). "Endosymbiotic adaptations in three new bacterial species associated with Dictyostelium discoideum: Paraburkholderia agricolaris sp. nov., Paraburkholderia hayleyella sp. nov., and Paraburkholderia bonniea sp. nov." PeerJ 8: e9151.

Brock, D. A., S. Read, A. Bozhchenko, D. C. Queller and J. E. Strassmann (2013). "Social amoeba farmers carry defensive symbionts to protect and privatize their crops." Nature Communications 4(1): 2385.

Gao, J., M. Liu, S. Shi, Y. Liu, Y. Duan, X. Lv, T. Bohu, Y. Li, Y. Hu and N. Wang (2021). "Disentangling Responses of the Subsurface Microbiome to Wetland Status and Implications for Indicating Ecosystem Functions." Microorganisms 9(2): 211.

Haselkorn, T. S., S. DiSalvo, J. W. Miller, U. Bashir, D. A. Brock, D. C. Queller and J. E. Strassmann (2019). "The specificity of Burkholderia symbionts in the social amoeba farming symbiosis: Prevalence, species, genetic and phenotypic diversity." Molecular ecology 28(4): 847-862.

Haselkorn, T. S., D. Jimenez, U. Bashir, E. Sallinger, D. C. Queller, J. E. Strassmann and S. DiSalvo (2021). "Novel Chlamydiae and Amoebophilus endosymbionts are prevalent in wild isolates of the model social amoeba Dictyostelium discoideum." Environ Microbiol Rep 13(5): 708-719.

Husnik, F. and P. J. Keeling (2019). "The fate of obligate endosymbionts: reduction, integration, or extinction." Curr Opin Genet Dev 58-59: 1-8.

Moran, N. A. (2003). "Tracing the evolution of gene loss in obligate bacterial symbionts." Curr Opin Microbiol 6(5): 512-518.

Noh, S., B. J. Capodanno, S. Xu, M. C. Hamilton, J. E. Strassmann and D. C. Queller (2022). "Comparative genomics support reduced-genome Paraburkholderia symbionts of Dictyostelium discoideum amoebas are ancestrally adapted professional symbionts." bioRxiv.

Parsons, M. and K. Fisher (2020). "Indigenous peoples and transformations in freshwater governance and management." Current Opinion in Environmental Sustainability 44: 124-139.

ResBaz Aotearoa: Building on Success

Tom Saunders, Laura Armstrong, Centre for eResearch, University of Auckland
Matt Plummer, Research Office, Victoria University of Wellington
Murray Cadzow, Mandy Phipps-Green, University of Otago
tom.saunders@auckland.ac.nz, l.armstrong@auckland.ac.nz, matt.plummer@vuw.ac.nz,
murray.cadzow@otago.ac.nz, mandy.phipps-green@otago.ac.nz

ABSTRACT

Researchers need access to the digital tools and skills at the heart of modern, high-impact research. Research Bazaar addresses this need by providing an opportunity to foster the exchange of ideas and make connections. Since 2016, several New Zealand universities have delivered ResBaz events. In 2022, the University of Otago joined the University of Auckland and Victoria University of Wellington in a collaborative approach to delivering the third online ResBaz Aotearoa event. Together, they have been able to offer a wide range of sessions to more of the research community than they could with their individual events.

This birds-of-a-feather session hopes to draw on the community's experience of in-person and online ResBaz and digital research skills offerings to explore how we can be more inclusive and sustainable. Questions for the session include:
- How do we attract more people to be involved as organisers or contributors?
- What can we do to appeal to the broadest range of attendees?
- What balance of in-person and online delivery supports attendee engagement and builds communities?
- How can we improve the logistics and administration of the event?

We will also invite participants to the session and other stakeholders to engage in a network to support ongoing collaboration.

This community consultation will guide the direction of a longer-term vision to upskill the New Zealand research community, and in particular, how the delivery of ResBaz events can contribute to this outcome.

ABOUT THE AUTHORS

Tom Saunders

Tom is an eResearch Engagement Specialist in the Centre for eResearch, University of Auckland. He has a background in entomology and biological control, but is now working to transfer digital research skills to doctoral students and researchers. He organises and co-teaches Carpentries workshops as a certified instructor, and is involved in organising and teaching in Research Bazaar Aotearoa. Tom maintains an interest in open research, and authored a report on increasing open access rates in Aotearoa during an internship in the Office of the Prime Minister's Chief Science Advisor ("The Future is Open").

Laura Armstrong

Laura is the eResearch Engagement Lead for Waipapa Taumata Rau | University of Auckland. She collaborates to engage with the research community to raise awareness and use of modern technologies and tools to advance research. Areas of focus include enabling researchers to

manage research data following best practices, including FAIR, CARE and Māori Data Sovereignty data principles, and providing digital research skills and community building.

Matt Plummer

Matt is a Senior Research Partner in Victoria University of Wellington's Research Office.With a background that spans the arts and technology, he works with researchers from different disciplines to facilitate collaborative projects, especially those which utilise technology in innovative and transformative ways. He's assisted with the delivery of a number of digital research training workshops and events, and the development of open source projects, augmented reality applications and research tools.

Murray Cadzow

Murray is a Scientific Programmer on the Research Teaching IT Support team at the University of Otago. Murray has a research background, and now works alongside researchers to help them achieve their computational research goals. In recent years he has been heavily involved in computational literacy and bioinformatic training at the University of Otago - organising ResBaz 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. Murray is a Genomics Aotearoa training associate, and a member of the NeSI Research Reference Group.

Mandy Phipps-Green

Mandy is Research Services Librarian at the University of Otago. Mandy has a background in complex disease genetics and forensic science, but has recently moved to a librarian role where she is part of a small team that works with researchers to exploit digital potentials, develop tools and services that enhance research outcomes, and to keep apprised of changes in the scholarly landscape. She manages the University of Otago's institutional repository, OUR Archive, and is the RSU/University of Otago Library's Carpentries liaison. She has attended many Otago ResBaz and Carpentry events, as a participant and as a helper, and recently undertook Carpentries instructor training. Mandy is particularly interested in open access in research, publishing practices, copyright, and the use of data science skills to support Library services and research.

Is your institution supporting FAIR research

Georgina Rae, Nooriyah Lohani, New Zealand eScience infrastructure
Paula Andrea Martinez, Australian Research Data Commons
georgina.rae@nesi.org.nz, nooriyah.lohani@nesi.org.nz, paula.martinez@ardc.edu.au

ABSTRACT

The FAIR guiding principles (Findable, Accessible, Interoperable and Reproducible) were originally developed for research data. However, software has fundamental characteristics that data do not possess. Hence there is a distinct need to explore how the FAIR principles can be applied to research software.

Work is now underway to create community-endorsed FAIR principles for research software and encourage their adoption via the ReSA/RDA/FORCE11 FAIR for Research Software (FAIR4RS) Working Group. In this BoF we will be bringing some of the recommendations and best practices of the practice group to the New Zealand eResearch audience who can help support researchers to develop good research software practices which will also contribute to sustainable development.

We propose a Birds of a feather session to raise awareness with institutions about what FAIR RS means globally as well as in a New Zealand context and discuss some key points institutions can take into consideration to enable FAIR research software development within their organisations. These key points range from tools that can help researchers to develop and follow good practices eg. Git, containers to the people who write and maintain research software (or research software engineers (RSE's)).

BoF Format: Lightning talks followed by a Panel discussion with participation from the audience.
Results: The outputs from this BoF will be published in the form of a blogpost highlighting tactics that institutions can apply to enable FAIR research software

REFRENCES
FAIR Principles https://www.go-fair.org/fair-principles/
TaskForces-ResearchSoftwareAlliance https://www.researchsoft.org/taskforces/

ABOUT THE AUTHORS

Nooriyah has a background in Genetics and Computer Science with 8 years experience working in commercial R&D and academic research in New Zealand. Nooriyah is now part of the engagement and communications team at NeSI alongside pursuing a PhD in Data Science. Nooriyah is passionate about recognising the contributions of researchers/research software engineers bringing together research and software skills, which she actively works towards as co-chair of the Research Software Engineers - Australia and New Zealand community.

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.

Paula Andrea Martinez is the Software Project Coordinator of the Australian Research Data Commons (ARDC) and Research Software Alliance (ReSA) Community Manager. She is leading projects of the Software Program to See, Shape and Sustain research software. She is a co-chair of the international FAIR4RS working group, co-host of the Visible Research Software interest group, and Steering Committee member of the Research Software Engineers RSE-AUNZ association. She has developed strategic planning and work packages to lead change in the recognition of research software as a first-class scholarly output of research

Building diverse teams through inclusive hiring practices

Jana Makar, New Zealand eScience Infrastructure (NeSI), jana.makar@nesi.org.nz Yesh Ramesh, REANNZ, yeshaswini.ramesh@reannz.co.nz
Yvette Wharton, Centre for eResearch (CeR), University of Auckland, y.wharton@auckland.ac.nz

ABSTRACT
Don't let unconscious bias deter prime candidates from applying to your roles! Understanding and adopting inclusive practices in your hiring process – from writing job descriptions to hosting interviews – will enable your organisation to draw from a wider pool of skills and talent, and build diverse teams that better represent the communities you serve.

In this Birds-of-a-Feather (BoF) session, we will kick off with 2-3 lightning talks that share some of the experiences, advice and approaches organisations in our sector have adopted to consider diversity and inclusion when building their teams. We'll then open the floor for questions and discussion, touching on topics such as:

- where / how are you recruiting
- who is on your hiring panel
- what questions are you asking in interviews
- how are you onboarding new team members
- why diverse teams make your organisation stronger

As an output from this session, a guidelines / recommendations document will be produced as a resource. Participants can then take this back to their organisations to put into practice right away or get conversations started about where and how positive changes can be made over time.

This session is supported by the Australasia Chapter of Women in High Performance Computing (WHPC+ AusNZ). We include a "+" in our name to emphasise that we welcome backgrounds and perspectives beyond gender to support greater diversity and inclusion across our HPC and eResearch sectors. To learn more about the Chapter and to connect with our community, visit https://aero.edu.au/whpc/

ABOUT THE AUTHORS
Jana Makar, Communications Manager, New Zealand eScience Infrastructure (NeSI) Based at the University of Auckland, Jana coordinates a variety of engagement and communications initiatives for NeSI. She's also led communications teams and projects in Canada with research & education networks, HPC consortia, and cybersecurity programs. She has a degree in Communications from the University of Calgary and spent the early part of her career working as a journalist in Western Canada.

Yesh Ramesh, Manager, Network Operations, REANNZ
Yesh manages Network Operations at REANNZ, Aotearoa New Zealand's national research and education network (NREN). She brings her extensive experience as a Senior Network Engineer, working earlier in her career with the IBM NZ team supporting the Vodafone IP Network. In India, she worked as a senior engineer at Juniper Networks, a Business Analyst at Capgemini and had various engineering roles at Alcatel-Lucent where she started her career. Since joining REANNZ in 2015, Yesh's work has involved leading the day-to-day operations of the network and project deliveries for REANNZ members.

Yvette Wharton, eResearch Solutions Lead, Centre for eResearch (CeR), University of Auckland
As eResearch Solutions Lead, Yvette works on research data management services and researcher enablement projects. She has extensive experience in University teaching, research and IT environments and is passionate about using her broad knowledge to facilitate people to achieve their aspirations.
http://orcid.org/0000-0002-6689-8840

Women in HPC - Australasia Chapter (WHPC+ AusNZ)
The Australasian Chapter of the global organisation Women in High Performance Computing (WHPC) is a collaboration between New Zealand eScience Infrastructure[1], NCI Australia[2], Pawsey Supercomputing Research Centre[3], Monash University[4], and Australasian eResearch Organisations (AeRO)[5]. The WHPC Organising Committee is composed of: Jana Makar[1], Adam Huttner-Koros[2], Aditi Subramanya[3], Marco de la Pierre[3], Kerri Wait[4], and Loretta Davis[5]. For more information, visit https://aero.edu.au/whpc/.

BoF - cloud use-cases in research

Blair Bethwaite, Jun Huh

NeSI

blair.bethwaite@nesi.org.nz, jun.huh@nesi.org.nz

ABSTRACT

It's been two years since our last cloudy BoF at eResearch 2021. Since then NeSI has been busy - one of the things we've done is fielded Flexible HPC (aka FlexiHPC), a new private cloud platform. We are now beginning to realise our vision of supporting use-cases that leverage FlexiHPC as a programmable infrastructure for research data collaboration. Whilst this is exciting, we need your input to help chart the course to wider benefit for the sector.

In this BoF we'll briefly touch on what FlexiHPC is and what we're doing with it (though other presentations will cover this in more detail). Our main focus will be to facilitate discussion with researchers, RSEs, eResearch support staff, and devops practitioners - so if this is you, please come along. We want to hear what you and your research communities would benefit from in the cloud services space.

This might include discussion around deploying specific virtual laboratory environments for NZ researchers and their collaborators, or suggestions/feedback on the sorts of shared services NeSI might offer to help you efficiently leverage cloud infrastructure like FlexiHPC (and beyond). Are there common tools you'd like to see available, what are the gaps and challenges, and things done on professor's personal servers that could benefit from wider access?

Bring your thinking caps and a willingness to share your experience in this space - there will be chocolate fish for participation!

Can we characterise Aotearoa New Zealand's research data at scale

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

ABSTRACT

This workshop seeks to translate the approach and findings from an Australian sector initiative on the characteristics of research data and its supporting infrastructures at scale, and catalyse a similar initiative in Aotearoa New Zealand. It will benefit individuals in senior management or senior operational staff of research institutions. In particular, those in senior advisory and planning roles in research data infrastructure, research data service owners, individuals from Research Offices and information management. Equally important is participation from those in research roles responsible for data governance and management at larger scales, where their role seeks to meet their wider research community and stakeholder needs.

The 'Macro View' of data developed by the Research Data Culture Conversation (www.researchdataculture.org) provides a first ever estimate of the scale of research data retained by Australian research institutions and national research infrastructure facilities for future access. Starting in 2017, this work built a consistent view across institutions, scaling from a few to a whole sector approach over time. The project team will describe how they carried this work forward, some of the insights gained, and how it is facilitating the conversations for change.

To bring this into an Aotearoa New Zealand context, prior to the workshop we will survey New Zealand Universities and Institutions to ask some of the key questions, such as "how much unique data does your institution hold?", and "given the importance of Indigenous Data Sovereignty in Aotearoa, and Te Tiriti requirements around active protection of taonga, what are the volumes and properties of Indigenous data in institutions?". This workshop will not only review the emerging 'Marco View' of Australian research data, but use first hand data from participating New Zealand Institutions alongside both Australian and New Zealand case studies to explore the process and findings so far. There will be a series of presentations from research institutes on activities underway that support infrastructure at scale.

The objectives of the session are to:
- Review this first ever estimate of Australia's retained research data;
- Characterise macro level trends in the types of data being retained;
- Define learnings and identify limitations for the macro view; and
- Receive input on defining next steps for Aotearoa New Zealand with regards to research data

Draft Agenda
- Introduction (Nick)
- The Australian view and lessons learnt (Rhys and Ai-Lin)
- What does it mean for New Zealand? (Nick/Claire)
- Lightning talks - institutional case studies
- Round table discussions (all)
- Closing conclusions

ABOUT THE AUTHORS

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). 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) and manages the Monash University response to the ARDC Data Retention and Institutional Underpinnings programmes.

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).

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 strategic directions and performance overall, 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 to establish a common framework for eResearch capability and investment nationally and institutionally.

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). 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).

'Enabling Māori, Pasifika, other minorities' Panel

Aim:
What & how are we and/or organisations doing positively in enabling Māori, Pasifika and other minorities to succeed in the research sector.

Format: Chair will introduce speakers - ask a question, allow audience to participate too in asking questions

Chair / Moderator: Rania Al-Ani - REANNZ

Speakers:

• Dr. Warren Williams, REANNZ Board Chair - http://www.karen.net.nz/about/our-people/dr-warren-williams/

• Te Taka Keegan - Professor & Researcher from UoW
https://www.waikato.ac.nz/news-opinion/media/2022/wananga-works-to-carve-a-pathway-for-maori-artificial-intelligence

ABSTRACT
What Are Minority Groups?
Sociologist Louis Wirth (1945) defined a minority group as "any group of people who, because of their physical or cultural characteristics, are singled out from the others in the society in which they live for differential and unequal treatment, and who therefore regard themselves as objects of collective discrimination."
According to Charles Wagley and Marvin Harris (1958), a minority group is distinguished by five characteristics: (1) unequal treatment and less power over their lives, (2) distinguishing physical or cultural traits like skin color or language, (3) involuntary membership in the group, (4) awareness of subordination, and (5) high rate of in-group marriage. Additional examples of minority groups might include the LGBTQ community, religious practitioners whose faith is not widely practiced where they live, and people with disabilities.

Relate to New Zealand – from the royal commission

Over the last 30 years the diversity of New Zealand's population has increased significantly in terms of ethnicity, culture, gender identities, religion, values, languages spoken, ages, sexual orientation and whānau structure. New Zealand has been described as a "superdiverse" country. Superdiversity means "a substantial increase in the diversity of ethnic, minority and immigrant groups in a city or country". One indicator of superdiversity is that a quarter of New Zealand's population was born overseas.

ABOUT THE AUTHORS

Dr Warren Williams has been in the ICT business and tertiary education industry for over 25 years. He is currently the CEO of 20/20 Trust that provides access and connection to digital skills and resources that empower and enable people to realise their aspirations.

His governance and leadership experience includes national and regional boards, working with Māori and entrepreneurship groups, tertiary ICT advisory groups, and technology and innovation groups. He is passionate about Strategic Māori leadership and succession planning in decision-making forums, especially with rangatahi Māori and wāhine Māori

Associate Professor Te Taka Keegan has worked on a number of projects involving the Māori language and technology. These include the Māori Niupepa Collection, Te Kete Ipurangi, the Microsoft keyboard, Microsoft Windows and Microsoft Office in Māori, Moodle in Māori, Google Web Search in Māori, the Māori macroniser and SwiftKey for Māori. In 2009 Te Taka spent 6 months with Google in Mountain View as a visiting scientist assisting with the Google Translator Toolkit for Māori. Further work with Google led to Translate in Māori. In 2013 Te Taka was awarded the University of Waikato's Māori/Indigenous Excellence Award for Research. In 2017 Te Taka was awarded the Prime Minister's Supreme Award for Tertiary Teaching Excellence.

An unexpected journey into PIDs - from planning to impact and back again

Yvette Wharton[1], Chris Seal[1], Natasha Simons[2], Jason Gush[3], Laura Armstrong[1], Mark Gahegan[1], Avishek Kumar[1]

[1]Waipapa Taumata Rau | University of Auckland,Tāmaki Makaurau Auckland, Aotearoa New Zealand
[2]Australian Research Data Commons, Brisbane, Australia
[3]Royal Society Te Apārangi, Aotearoa, New Zealand
y.wharton@auckland.ac.nz, c.seal@auckland.ac.nz, natasha.simons@ardc.edu.au,
l.armstrong@auckland.ac.nz, m.gahegan@auckland.ac.nz, a.kumar@auckland.ac.nz

ABSTRACT

Join our quest to reclaim the lost connections in research from the clutches of the dragon of chaos. Our wizards and hobbits leading today's quest will guide us through the wild forests of Murky data, towards the lost mountains of disorganised and duplicated data. Through perseverance and the occasional trail song, we will gather the tools needed to organise and connect lost artefacts, hidden in the dragon's treasure hoard, bringing the golden rewards of more FAIR data, governed by CARE principles.

Along the way we will meet with a fellowship of persistent identifiers (PIDs), knowledge tags and data management plans, and together we will embark on a journey from planning to impact and back again.

Grab your backpack, enjoy second breakfast (and elevensies) and join us to explore our fellowship of PIDs and discuss strategies to maximise benefits and guide our journey through treacherous goblin-filled lands including:

- Where could you re-use existing metadata?
1. What actions can be triggered or automated with the presence of a PID, and should we do it?
2. What aggregation and analysis of entities at an institutional or national scale could we do, and what are we seeking to uncover?

ABOUT THE AUTHORS

Chris Seal: https://orcid.org/0000-0002-4850-1395
Natasha Simons:https://orcid.org/0000-0003-0635-1998
Jason Gush: https://orcid.org/0000-0001-8920-0452
Laura Armstrong: https://orcid.org/0000-0003-2370-3924
Mark Gahegan: https://orcid.org/0000-0001-7209-8156
Avishek Kumar: https://orcid.org/0000-0003-0069-8122
Yvette Wharton: https://orcid.org/0000-0002-6689-8840

Working with Singularity Containers (Interactive Workshop)

Wolfgang Hayek, Callum Walley, Chris Scott, Dinindu Senanayake, Maxime Rio
NIWA, NeSI
wolfgang.hayek@niwa.co.nz, callum.walley@nesi.org.nz, chris.scott@nesi.org.nz,
dinindu.senanayake@nesi.org.nz, maxime.rio@nesi.org.nz

ABSTRACT / INTRODUCTION

Containers have become an indispensable part of many workflows in modern research computing, supporting software portability and reproducibility of results, and allowing software developers to bundle and distribute codes with complex dependencies as a ready-to-go solution. Various container runtimes have been created over the years, including the well-known Docker and Singularity packages. Singularity is widely used in many research computing disciplines and in high-performance computing, as its design supports these workflows well and integrates with accelerators and the Message Passing Interface (MPI) for distributed parallel computation.
This half-day hands-on workshop covers essential aspects of using Singularity, including the basics of containerisation, how to run and interact with containers, and how to create new containers. The workshop will be conducted on NeSI's Mahuika HPC, so attendees should bring a laptop, have a NeSI account, and should have basic command line skills. They are free to run Singularity on their own Linux system if they prefer, provided that they have sufficient privileges to install the software and run containers. Further instructions will be sent out ahead of the workshop.
[Note to the conference organisers: this workshop could also be offered as a hybrid workshop, or fully virtual. In the fully virtual case, the workshop could be run as a pre-conference event. In any case, attendees should be asked to register, as we will need to cap attendance at around 30 people. All authors of this abstract will instruct or help at the workshop]

ABOUT THE AUTHORS

Wolfgang Hayek is a HPC Research Software Engineer at NIWA, and group manager of NIWA's scientific programming group, with many years of experience in scientific computing and HPC.

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.

Chris Scott is a NeSI Research Software Engineer based at The University of Auckland.

Building a taonga species data repository for Aotearoa New Zealand

Mik Black, University of Otago
Libby Liggins, School of Natural Sciences, Massey University
Rudiger Brauning, AgResearch
Jun Huh, Claire Rye, Eirian Perkins, New Zealand eScience Infrastructure
mik.black@otago.ac.nz, L.Liggins@massey.ac.nz, rudiger.brauning@agresearch.co.nz, jun.huh@nesi.org.nz, claire.rye@nesi.org.nz, eirian.perkins@nesi.org.nz

ABSTRACT

The Aotearoa Genomic Data Repository (AGDR) is a joint initiative between Genomics Aotearoa and NeSI to provide a secure NZ-based option for the storage, management and sharing of non-human genomic data generated from biological and environmental samples originating in Aotearoa New Zealand. This resource has been developed to follow the principles of Māori Data Sovereignty, and to enable the right of kaitiakitanga, so that iwi, hapū and whānau can effectively exercise their responsibilities as guardians over species that they regard as taonga. While the repository is designed to facilitate the sharing of data - making it findable by researchers, and interoperable with data held in other genomic repositories - the decision-making process regarding who can access the data is entirely in the hands of those holding kaitiakitanga over each data set. No data are made available to the requesting researcher until the request has been approved, and the conditions for access (which can vary by data set) have been agreed to. This talk will provide an update on the progress made within the AGDR project over the past year, including the incorporation of Traditional Knowledge and Biocultural Labels and Notices, and digital object identifier minting for data sets, as well as providing an overview of the process for submitting genomic data sets to the repository.

ABOUT THE AUTHORS

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

Libby Liggins is an Associate Professor at Massey University, and a Research Associate of the Auckland Museum Tāmaki Paenga Hira in Aotearoa New Zealand. She specializes in using genomic data to examine the generation, maintenance, and erosion of marine biodiversity. Libby participates in several initiatives interested in the ethical stewardship and aggregation of data. She is on the Scientific Committee for the International Science Council's World Data System, and the Steering Committee for the Genomic Observatories Metadatabase (GEOME) and the Group on Earth Observations Biodiversity Observation Network (GEOBON) Genetic Composition Working Group.

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

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).

Rakeiora a pathfinder for genomic medicine: Building a prototype Genomics Platform

Huti Watson, Ngāti Porou Hauora, huti.watson@nph.org.nz
Jun Huh, New Zealand eScience Infrastructure, jun.huh@nesi.org.nz
Claire Rye, New Zealand eScience Infrastructure, claire.rye@nesi.org.nz
Cristin Print, University of Auckland, c.print@auckland.ac.nz
Stephen Robertson, University of Otago, stephen.robertson@otago.ac.nz
Muhammad Qasim, Ngati Porou Hauora, muhammad.qasim@nph.org.nz
Phillip Wilcox, University of Otago, phillip.wilcox@otago.ac.nz
Ben Te Aika, University of Otago, ben.teaika@otago.ac.nz
Helen Wihongi, Waitemata District Health Board, helen.wihongi@waitematadhb.govt.nz
Donia Macartney-Coxson,Institute of Environmental Science and Research (ESR), Donia.Macartney-Coxson@esr.cri.nz
Joep de Ligt, Institute of Environmental Science and Research (ESR), Joep.DeLigt@esr.cri.nz
Nathalie Giraudon, New Zealand eScience Infrastructure, nathalie.giraudon@nesi.org.nz
Matt Pestle, New Zealand eScience Infrastructure, matt.pestle@nesi.org.nz
Eirian Perkins, New Zealand eScience Infrastructure, eirian.perkins@nesi.org.nz
Kenny Zhao, New Zealand eScience Infrastructure, kenny.zhao@nesi.org.nz
Ben Curran, University of Auckland, b.curran@auckland.ac.nz

ABSTRACT
The Rakeiora project is a 'pathfinder' to prototype how best to acquire, protect, use and store genomic datasets for analysis support of primary and tertiary health research in Aotearoa New Zealand. This work is at heart about establishing an innovative approach which embraces tikanga Māori and supports community-led approaches and needs for Māori Data Sovereignty. This approach is being applied to a specific setting, genomics research in primary and tertiary health care settings in Aotearoa New Zealand, but much of the methodology is applicable to embracing and meeting the expectations of indigenous people in any community.

The tikanga attributes which will be discussed, were developed early on and have been a crucial part of the project and these underpin the co-design approach that has been taken. There has been a genuine desire to take on board and incorporate the ideas and perspectives of not just the researchers who will use the system, but of the kaitiaki, to ensure data sovereignty is respected and upheld while facilitating appropriate research and improved outcomes to the individuals who are participating.

As the Rakeiora project reaches the end of its prototype phase, the team would like to reflect on the journey thus far, the design decisions, lessons learnt and demo the platform as it is now. The demo will highlight how the platform attempts to deliver on the tikanga attributes walking through the user journey of a researcher applying for access, the access being granted and then running a workflow. We will then show the behind the scenes but equally important kaitiaki approval (and rejection) process and the audit functionality of the system.

Journeying towards good data management practices: Tips and tricks to empower the biodiversity genomics community

Jana Wold*[1,2], Natalie Forsdick*[3,2], Anton Angelo[4], François Bissey[5], Jamie Hart[5], Mitchell Head[6], Libby Liggins[7], Dinindu Senanayake[8], Tammy Steeves[1]
* Co-presenters: jana.wold@canterbury.ac.nz & forsdickn@landcareresearch.co.nz
Organisations:
1 School of Biological Sciences, University of Canterbury
2 Genomics Aotearoa
3 Manaaki Whenua – Landcare Research; Genomics Aotearoa
4 Library, University of Canterbury
5 Digital Services, University of Canterbury
6 Te Kotahi Research Institute, University of Waikato
7 School of Natural Sciences, Massey University
8 New Zealand eScience Infrastructure

ABSTRACT

Large-scale genomic data sets are being generated at an unprecedented pace, requiring collaboration across disciplines to utilise these data to their full potential. We represent a cross-disciplinary team, including IT and eResearch support staff, biodiversity genomics researchers, and those interested in appropriate data stewardship and re-use. Through our experiences, we appreciate the need for appropriate data management practices not only to minimise the risks of data loss and maximise efficiency and productivity for single projects, but also to enhance opportunities for future data re-use and collaboration. These practices are an essential component in facilitating Indigenous data sovereignty and are an expectation of the FAIR and CARE principles. Challenges associated with biodiversity genomic data sets include increasing quantity of data, variability in computational platform, changing data storage requirements, and the cross-institutional and -generational nature of research projects.

To address these challenges, we discuss three case studies based on personal journeys in the data management space. We use these examples to demonstrate the realistic considerations, compromises, and actions for biodiversity genomic data management. These include but are not limited to project life-stage, compute platform, data characteristics, and the aspirations of Indigenous communities. We then present tips and tricks to support emerging researchers (and research teams more broadly) in documenting and managing biodiversity genomic data sets and associated metadata. These include the provision of a GitHub repository that models good data management practices for dissemination across the research community. We aim to empower the biodiversity genomics community to minimise risks and maximise research impact now and into the future.

ABOUT THE AUTHORS

Jana Wold and Natalie Forsdick are postdoctoral research fellows who leverage genomic data to inform conservation management of threatened species in Aotearoa. They are particularly interested in incorporating evolutionary concepts to population-scale management. Jana works to incorporate novel computational methods into conservation genomics including genome graphs. Natalie's research centres around building genomic resources for a broad array of taxa.

The first RSE Asia Australia Unconference

Authors: Paula Andrea Martinez, Australian Research Data Commons
Rowland Mosbergen, WEHI
Saranjeet Kaur Bhogal, RSE Asia Association
Nooriyah Lohani, New Zealand eScience Infrastructure
paula.martinez@ardc.edu.au, mosbergen.r@wehi.edu.au, kaur.saranjeet3@gmail.com,
nooriyah.lohani@nesi.org.nz

ABSTRACT
The RSE AUNZ (rse-aunz.org) Association started in 2018. Growing a community that promotes the value of Research Software Engineers in Australia and New Zealand, thanks to grassroots efforts. In 2022, we ran the first RSE Asia Australia unconference online with the theme "Building community" in joint partnership between the newly formed RSE Asia Association. We ran the unconference during September, the month of RSE conferences around the world, aligning with the RSECon2022, in the UK and the online NZRSE 2022. The unconference mode ensured the members of our community had the space to share their valuable technical and social knowledge of research software with the community**.**

Before the unconference we gained support through our partners (ARDC, QCIF, Society for RSE, NCI Australia and Google Cloud for researchers) which ensured access to scholarships and low registration fees. In total, we attracted 141 registrations from 10 countries reaching our goals of spread and inclusion. We facilitated 3 time slots which prompted 17 community-led discussions. We also prearranged 3 keynote speakers and a panel on career progression. The maximum attendance was 82 people in one session.

In retrospect, the feedback was outstanding, with many people acknowledging the opportunity to connect with members of the community and the high level of flexibility of the unconference. We also received very favourable feedback from partners, with a commitment of support continuation and opportunity for new partners to come on board. From the feedback survey, 85% of feedback respondents said they will recommend the RSE AUNZ to their colleagues which demonstrates the impact the RSE unconference produced within attendees. The RSE AUNZ Association is in favour of running a similar event jointly organised with the NZ RSE Conference in the coming years.

ABOUT THE AUTHORS
Dr Paula Andrea Martinez is the Software Project Coordinator at the Australian Research Data Commons (ARDC). She leads projects that improve the recognition of research software and research software authors. She is also an advocate of open science and diversity and active in various community projects such as the Research Software Alliance (ReSA), the Visible Research Software Interest Group (VRS), the Association of Research Software Engineers in Australia and New Zealand (RSE AUNZ). She has a background in Computer Science and Bioinformatics, with 16 years of experience in industry and academia.

Rowland Mosbergen is a "translator" with 22 years experience in it as a strategic advisor, developer, analyst/programmer, team leader, software architect and small business owner. He has 12 years of experience in research, 8 of those with the Stemformatics project - a bioinformatics tool that allows biologists to visualise their gene expression data online. Rowland has worked at the corporate level with large financial institutions like Merrill Lynch and National Australia Bank on their risk management systems and at the small business level to

provide computer support. He is now working as a strategic advisor in the Research Computing Platform - an RSE group that works within WEHI.

Nooriyah has a background in Genetics and Computer Science with 8 years experience working in commercial R&D and academic research in New Zealand. Nooriyah is now part of the engagement and communications team at NeSI alongside pursuing a PhD in Data Science. Nooriyah is passionate about recognising the contributions of researchers/research software engineers bringing together research and software skills, which she actively works towards as co-chair of the Research Software Engineers - Australia and New Zealand community.

Saranjeet is a Statistician by training, based in India (Asia) and an Open Source and Open Science Advocate. She is the Co-Founder of the RSE Asia Association and has been awarded the Research Software Engineering (RSE) Impact Award 2022, at the inaugural Community Awards by the Society of Research Software Engineering for her work on increasing the visibility and diversity of the RSE community.

Improving Trust in Identity Federations: an Update on New Developments in Tuakiri

Vladimir Mencl

Research and Education Advanced Network New Zealand

vladimir.mencl@reannz.co.nz

ABSTRACT

Tuakiri, the New Zealand Access Federation, has been serving the New Zealand R&E community for over a decade. As the landscape in the global R&E Identity and Access Management community that Tuakiri connects to keeps evolving, Tuakiri has to adopt these changes as well. This talk will provide an update on these changes, specifically:

- Changes in user-identifier attributes: adopting the new samlSubjectID and samlPairwiseID attributes, moving away from eduPersonPrincipalName and eduPersonTargetedID
- Changes in the way of expressing assurance levels, adopting the new REFEDS Assurance Framework, moving away from the current Level of Assurance values based on the NIST-SP-800-63 standards family.

The talk will also present the Tuakiri OpenID Connect Bridge, a new service allowing OpenID Connect-enabled services to use Tuakiri for authentication. The OpenID Connect bridge supersedes the RapidConnect service and new services should use OpenID Connect – but RapidConnect will stay operational to support existing services.

The talk will also cover other new developments in Tuakiri, incl. exposing Tuakiri usage stats in the REANNZ Member Portal.

ABOUT THE AUTHOR

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

Flexi HPC - Enabling new ways of eResearch

Blair Bethwaite, Thomas Berger, Jun Huh
New Zealand eScience Infrastructure (NeSI)
blair.bethwaite@nesi.org.nz, thomas.berger@nesi.org.nz, jun.huh@nesi.org.nz

ABSTRACT
In 2021, the New Zealand eScience Infrastructure (NeSI) procured and integrated FlexiHPC, 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 early use-cases for the new infrastructure by walking through the initial 3+ tenants, highlighting key partnerships and integrations.
The FlexiHPC platform is being designed to support multiple models of consumption and integration, including contemporary on-demand access to shared services alongside dedicated hardware tenancies with deep local integration. FlexiHPC aspires to play a core role in helping RSE/DevOps users to build and grow their research ecosystem and share across the research communities.
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.
NeSI is taking the approach of supporting multiple types of tenancy models to address the different needs of partnering organisations. This talk will also touch upon recent developments on the FlexiHPC, with the experience of strong partnership with AgResearch in particular, showcasing different types of tenancy models that can include co-ownership of infrastructure and shared responsibility model.

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

Automated CT analysis to identify low emitting ruminants

Dr Sam Hitchman, Laura Grundy, Dr Marlon Reis
Agresearch ltd
Sam.hitchman@agresearch.co.nz

ABSTRACT
Genetic selection for reduced methane production has been recognised by government as an important tool within the agricultural industry's toolbox in achieving government targets for reduced greenhouse gas emissions. Recent research has shown that breeding for lowered methane production is associated with changes in rumen physiology (size, constituents) and feed efficiency. It was found that rumen volume was significantly lower in low methane producing animals, within methane selection line animals which are divergently selected for methane yield [1].

Rumen volume and composition can be measured using computed tomography (CT) scanning of small ruminants, such as sheep and deer, whereby the entire animal is scanned at 5mm intervals. CT scanning of breeding animals is already implemented in these industries as part of a suite of imaging technologies to generate breeding values associated with animal [2]. To date CT image analysis has been used to predict carcass composition, such as lean meat yield [3]. These scans also collect excellent data of the other body constituents such as the rumen, with high contrast between gas, liquid, and raft components. However, manual processing of the CT images to estimate rumen volume and composition is a major hurdle for industry uptake. There is a need to automate this process to facilitate further research into low methane producing animals and reduce costs. Due to the complex boundaries of the rumen and its constituents, as well as the large variations between animals, traditional imaging processing could not be implemented for automatic segmentation and analysis of the rumen in live animals.

Automatic semantic (pixel wise) segmentation of greyscale images (such as those collected by CT images) has been possible in recent years with the development of deep learning techniques for computer aided diagnosis. Many applications involving image classification and segmentation have been developed, including tumour segmentation in mammogram images [4], classification of diseased brains [5], and automated segmentation for abdominal composition [6]. Using a dataset of 223 manually processed full body CT scans of sheep and deer previously collected by AgResearch, Deep Learning techniques to segment the rumen from the surrounding tissues and analyse rumen size and composition has been investigated, facilitated by the hardware available at the New Zealand eScience Infrastructure. Latest models perform with DICE scores exceeding .90 on an independent test set, showing that these techniques could soon provide unprecedented data for industry and research.

References

[1] Bain, W., Bezuidenhout, L., Jopson, N., Pinares-Patiño, C., & McEwan, J. (2014). Rumen differences between sheep identified as being low or high methane emitters. Paper presented at the 10th World Congress on Genetics Applied to Livestock.
[2] Scholz, A. M., Bünger, L., Kongsro, J., Baulain, U., & Mitchell, A. D. (2015). Non-invasive methods for the determination of body and carcass composition in livestock: dual-energy X-ray absorptiometry, computed tomography, magnetic resonance imaging and ultrasound: invited review. Animal, 9(7), 1250-1264. doi: 10.1017/S1751731115000336

[3] Gardner, G. E., Williams, A., Siddell, J., Ball, A. J., Mortimer, S., Jacob, R. H., Pearce, K. L., Hocking Edwards, J. E., Rowe, J. B., & Pethick, D. W. (2010). Using Australian Sheep Breeding Values to increase lean meat yield percentage. Animal Production Science, 50(12), 1098-1106. doi: https://doi.org/10.1071/AN10144

[4] Singh, V. K., Rashwan, H. A., Romani, S., Akram, F., Pandey, N., Sarker, M. M. K., Saleh, A., Arenas, M., Arquez, M., & Puig, D. (2020). Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network. Expert Systems with Applications, 139, 112855.

[5] Gao, X. W., Hui, R., & Tian, Z. (2017). Classification of CT brain images based on deep learning networks. Computer Methods and Programs in Biomedicine, 138, 49-56. doi: https://doi.org/10.1016/j.cmpb.2016.10.007

[6] Weston, A. D., Korfiatis, P., Kline, T. L., Philbrick, K. A., Kostandy, P., Sakinis, T., Sugimoto, M., Takahashi, N., & Erickson, B. J. (2019). Automated Abdominal Segmentation of CT Scans for Body Composition Analysis Using Deep Learning. Radiology, 290(3), 669-679. doi: 10.1148/radiol.2018181432

Testing Spatial Navigation with Google Street View

Murray Cadzow, and David Bilkey

University of Otago

murray.cadzow@otago.ac.nz, david.bilkey@otago.ac.nz

ABSTRACT / INTRODUCTION

A 2nd year psychology teaching lab was developed to allow students to test hypotheses concerning spatial navigation in an interactive virtual environment. The lab was implemented using Google Street View to take advantage of the wealth of environments available and involved a series of experiments to test two hypotheses. The first was that distance perception will be modulated by stride length, as predicted by theories of embodied cognition. The second was that navigation errors will compound with path integration in virtual space as observed in real-world navigation. Experiment 1 involved participants moving in a straight line at an unfamiliar Street View location and estimating when they had travelled 100m. Participants also had their average stride length measured. It was predicted that participants with a longer stride would perceive the distance as shorter due to their real-world experience where fewer strides needed to be taken to cover a set distance. In Experiment 2, participants navigated through narrow streets and paths, relying on local cues for navigation. They periodically estimated the distance and bearing back to the origin. It was predicted that as participants undertook their journey, errors in estimated distance and bearing would increase, as is known to occur with path integration in real-world settings.

The results from experiment 1 did not support a relationship between stride length and distance estimation. The results from experiment 2 showed an increase in estimation error occurred with greater distance along the journey, which supported the second hypothesis. Overall, this project illustrates how the massive amount of spatial data available in Google Maps and Street View can be used to set up interesting experiments that explore spatial navigation in a virtual environment. This talk will discuss some of the benefits and challenges of using this resource to develop research and teaching material.

ABOUT THE AUTHOR(S)
- Dr Murray Cadzow

Murray is a Scientific Programmer on the Research Teaching IT Support team at the University of Otago. Murray has a research background, and now works alongside researchers to help them achieve their computational research goals. In recent years he has been heavily involved in computational literacy and bioinformatic training at the University of Otago - organising ResBaz 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. Murray is a Genomics Aotearoa training associate, and a member of the NeSI Research Reference Group.

- Professor David Bilkey

David is a professor in the Psychology Department at the University of Otago. He is interested in systems neuroscience, with a particular focus on the role of the temporal cortex and prefrontal regions of the brain in memory and learning processes, and how this might be dysfunctional in schizophrenia. He also teaches at 1st and 2nd year level and coordinates 2nd year Psychology papers. He is interested in developing the associated laboratory classes to better suit computer-literate students.

Getting to know what runs in my private research cloud

Dr François Bissey
University of Canterbury
francois.bissey@canterbury.ac.nz

ABSTRACT

When UC started to offer a private cloud for research (RCC), I knew that I would have to track what was going on the platform. This presentation is about my journey of tracking and figuring how much it is used, what is used and how it is used. From humble beginnings with spreadsheets, google docs and a trello board to a MariaDB database with python tools for administration and queries and python dashboards to report and inform decisions.
In the end, it turned out that with the right skills, it is quite easy to build a powerful DIY framework of reporting with just a few open-source packages.

ABOUT THE AUTHOR

François is a particle physicist that knows too much about computers. On that basis he joined BlueFern at university of Canterbury at the end 2010 to be one of the people who would form the core of the first iteration of NeSI. When UC left NeSI he stayed at UC to help with what we called then the long tail of research computing which lead him to be managing the UC research cloud, mostly solo, for the past 5 years.

Incentivising Open Research

John Creech1, Michelle Blake2, Tom Saunders3, Laura Armstrong3, Mark Gahegan4, Chris Seal3, Mike
Laverick3, Jess Howie2, Gene Soudlenkov5
1Ministry of Business, Innovation and Employment (MBIE)
2 University of Waikato
3 Centre for eResearch, University of Auckland
4 Beyond Prediction, MBIE-funded Data Science Platform
5 National eScience Infrastructure

John.Creech@mbie.govt.nz, michelleee.blake@waikato.ac.nz, tom.saunders@auckland.ac.nz,
l.armstrong@auckland.ac.nz, m.gahegan@auckland.ac.nz, c.seal@auckland.ac.nz,
mike.laverick@auckland.ac.nz, jessica.howie@waikato.ac.nz, gene.soudlenkov@nesi.org.nz

ABSTRACT / INTRODUCTION

The international research community is moving forward with open research with an increasing
focus on valuing a diversity of research outputs, broadening the scope of research assessment
and measures of societal impact, and incentivising cross-disciplinarity.

In this BoF session we will explore different approaches to encouraging open research within the
Aotearoa New Zealand research sector. Research funders and institutions developing open
research/access/data policies and incentives are invited to share the steps they have taken or
hope to take to navigate the opportunities and challenges of open research.

Attendees will hear an update on MBIE's work on Open Research, and from those working within
institutions to enable different parts (e.g. code, software, instrument data, datasets, and outputs)
of the research workflow to be open, FAIR, and reproducible.

This will be a primer to group discussion focusing on:

-        What can funders and institutions do to incentivise Open Research behaviours?
-        What are the barriers to wider adoption of Open Research practices?
-        What kind of training and support do researchers need to understand and apply Open
Research principles?
Researchers, librarians, research support, policy makers and eresearch service providers are
invited to share their perspectives and practical advice on enabling Open Research.

ABOUT THE AUTHOR(S)
John Creech
John is a Senior Policy Advisor in the Science Policy team at MBIE. His work focus spans the data side of the
RSI system, including Open Science, Māori data sovereignty, research metrics, impact of research, and
research infrastructure. John has a background in planetary science, having previously worked in isotope
geochemistry and cosmochemistry research in New Zealand, Europe and Australia.

Michelle Blake
Michelle is the University Librarian at Te Whare Wānanga o Waikato, the University of Waikato. She previously worked in the UK for 15 years including as the Deputy and then Director of Library and Archives at the University of York. Michelle is a member of the Open Access Australasia Executive Committee, a founding member of the Committee on Data in Research (CoDiR) in Aotearoa, NZ, a member of CAUL's Indigenous Knowledge Advisory Group, a member of the International Association of University Libraries (IATUL) Board and co-leads the Open Scholarship strand of the CONZUL strategy.

Tom Saunders
Tom is an eResearch Engagement Specialist in the Centre for eResearch, University of Auckland. He has a background in entomology and biological control, but is now working to transfer digital research skills to doctoral students and researchers. He organises and
co-teaches Carpentries workshops as a certified instructor, and is involved in organising and teaching in Research Bazaar Aotearoa. Tom maintains an interest in open research, and authored a report on increasing open access rates in Aotearoa during an internship in the Office of the Prime Minister's Chief Science Advisor ("The Future is Open").

Laura Armstrong
Laura is the eResearch Engagement Lead for Waipapa Taumata Rau | University of Auckland. She collaborates to engage with the research community to raise awareness and use of modern technologies and tools to advance research. Areas of focus include enabling researchers to manage research data following best practices, including FAIR, CARE and Māori Data Sovereignty data principles, and providing digital research skills and community building.

Mark Gahegan
Mark is the director for the Centre for eResearch and a professor of Computer Science at Waipapa Taumata Rau | University of Auckland. His research interests are in eScience and Information Science. He is PI of one of the MBIE-funded National Data Science Platforms entitled "Beyond Prediction: explanatory and transparent Data Science", which aims to create new methods for opening up scientific workflows.

Chris Seal
Chris is a Senior eResearch Solutions Specialist for Waipapa Taumata Rau | University of Auckland. He has a background in Materials Science with a focus on fracture of metals and is now working in the Centre for eResearch developing an instrument data repository. A dedicated 'Persistent Identifiers (PIDs) enthusiast', he is a member of the I4IOz community of practice, developing best practice for the new Instrument PIDs, and is a member of the Research Activity ID (RAiD) advisory group.

Mike Laverick
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.

Jess Howie
Jess Howie is the Open Research Manager at the Te Whare Wānanga o Waikato, the University of Waikato and leads the Aotearoa Scholarly Communication Community of Practice. She has been involved in the CONZUL Open Access Project since 2019. This project gathers data on Open Access progress in Aotearoa annually.

Gene Soudlenkov
Gene is a Reseacher Support Team Lead for New Zealand eScience Infrastructure. Gene's background is in applied mathematics and software development. He specialises on C/C++/Fortran development and optimization, parallel programming and data analysis.

eresearchnz2023.org.nz
#eresearchnz2023