AeRO Forum 2024

26 July 2024 | 9:30am - 12:30pm
La Trobe University, Bundoora VIC

AeRO
Australasian eResearch Organisations

Research data movement and orchestration

The Australasian eResearch sector is continually evolving and increasingly complex.

A unique opportunity to explore best practices and innovative solutions for optimising data management in research environments.

Innovative Data Solutions	Research Data Movement Panel	eResearch Networking

9:00am	**Networking Coffee**
9:30am	**Introductions:** *Luc Betbeder-Matibet, UNSW*
9:40am	**Data movement in national digital research infrastructure** *Prof David Abramson, University of Queensland*
9:50am	**Institutional data fabrics – time to go wider. Liberate, collaborate, succeed.** *Jake Carroll, University of Queensland*
10:10am	**Microscopy in the big-data era: a journey for better data movement** *Dr David Poger, Microscopy Australia*
10:30am	**Morning tea**
11:00am	**Moving research data in the geosciences** *Dr Rebecca Farrington, AuScope*
11:10am	**Current status and challenges in accessing data from international repositories: An overview for Australian life scientists** *Dr Farah Zaib Khan, Australian BioCommons*
11:20am	**Automatically acquiring instrument data & relocating repositories** *Craig Hamilton, Intersect*
11:30am	**Managing big data from a fleet of mobile laboratories** *Marcus John, National Transport Research Organisation (NTRO)*
11:40am	**Panel discussion – addressing growing research demands:** *Moderator: Gareth Williams* *Jake Carroll, Dr David Poger, Dr Rebecca Farrington, Dr Farah Zaib Khan*
12:20pm	**Wrapping up – collaboration for evolving data challenges:** *Greg Darcy, AARNet*
12:30pm	**Forum end**

ADVOCATE   **COLLABORATE**   **COMMUNICATE**   **aero.edu.au**

9:40am | *Data movement in national digital research infrastructure*
Prof David Abramson, Professor of Computer Science, University of Queensland

As research workflows expand to encompass multiple computing and storage environments, could data movement be the glue that binds a globally networked and coordinated national digital research infrastructure (NDRI)?

9:50am | *Institutional data fabrics – time to go wider. Liberate, collaborate, succeed.*
Jake Carroll, Director, Research Computing Centre, University of Queensland

In this brief talk, Jake will explain the readiness of our sector to move to the next wave of research data storage principles, focusing on data fabrics, data movement and the fundamental rational behind the next generation of architecture. Data fabrics have become popular, relatively easy to deploy and offer compelling and robust approaches to sharing, economies of scale, management, and governance. Jake will highlight real world research use cases in cryogenic transmission electron microscopy (CryoTEM), genomics, drug discovery, and distributed artificial intelligence model training in production right now between institutions. He will explain how we've addressed the significant challenges that came before, and where we're going to take this next, as these architectures, design patterns and strategies continue to prove themselves viable at extreme scale.

About the presenter: Jake Carroll is the Director of The University of Queensland Research Computing Centre. Over his 19+ year career he has helped shape the local, state and national agenda on eResearch infrastructure, empowering the sector and supporting some of the most vibrant scientific discovery efforts in our nation. Jake blends academia, scientific method, empathy, and innovation to lead a team of dedicated and focused people, passionate and motivated by value-co-creation with their clients. Jake is an elected member of the International Storage Board Industry Advisory Council and is a UQ Business School MBA Alumnus.

10:10am | *Microscopy in the big-data era: a journey for better data movement*
Dr David Poger, Research Data Manager, Microscopy Australia

Over the past few years, considerable technological progress in instrumentation and software tools in microscopy has paved the way for unprecedented scientific advances. This has resulted in ever-increasing volumes of data produced at high speed and resolution. This has, however, also posed substantial challenges to microscopy facilities in terms of workflow optimisation and data management. For example, moving data from temporary storage at the facilities to their users' storage systems (workstations, institutional storage etc) has often become more difficult than it used to be. While widely used by facilities and researchers for many years, traditional tools such as FileZilla, CuteFTP, Cyberduck and WinSCP too often lead to unsuccessful transfer of large data sets, causing frustration and delays in clearing temporary storage at facilities.

As part of the Australian Characterisation Commons at Scale (ACCS) project, Microscopy Australia promoted and coordinated a testing phase of the data-transport service Globus at microscopy facilities in partnership with AARNet. In addition to fast and secure data transfer, Globus enables reliable transfer of large data volumes regardless of the distance between endpoints. Furthermore, Globus can support data workflow automation by scheduling data transfer. Thanks to its reliability and integration into workflows, Globus has greatly improved data transfer and sharing at the facilities that have adopted Globus. Since the end of the ACCS project in June 2022, a growing number of universities and microscopy facilities have adopted or planning to adopt Globus.

About the presenter: Dr David Poger was awarded his PhD in Structural Biology from Joseph Fourier University (Grenoble, France) in 2005. He then moved to Australia where he worked as a Research Fellow at The University of Queensland, studying the structure and dynamics of biological membranes, protein receptors and membrane-active compounds using computer simulations. In 2020, David joined Microscopy Australia, a consortium of nine university-based microscopy facilities, as Research Data Manager. He assists microscopy facilities in data management by developing best practices and guiding them in their journey to FAIR data. David participates actively in several projects and working groups in Australia and overseas to develop and promote best practices in research data management and facilitate the interaction between microscopy facilities, data managers and IT/eResearch specialists.

11:00am | *Moving research data in the geosciences*
Dr Rebecca Farrington, Director of Research Data Systems, AuScope

As Australia's provider of research infrastructure to the geoscience community, AuScope acquires, processes, curates, preserves, and reuses research data from fields, laboratories, and computational programs in many geoscience subdomains, including the geophysics, geochemistry, and geospatial sciences. This talk will present case studies and highlight key challenges from the national geoscience research community.

About the presenter: Rebecca is the Director of Research Data Systems at AuScope Ltd, Australia's national provider of Research Infrastructure for the Geoscience community. Prior to this, she was a Senior Research Fellow within the School of Geography Earth and Atmospheric Sciences and an Academic Convenor on the Petascale Campus Initiative (Chancellery) at The University of Melbourne. With research expertise in computational fluid dynamics, she is passionate about developing community-led data and compute-intensive research initiatives and supporting the people behind them.

11:10am | *Current Status and Challenges in Accessing Data from International Repositories: An Overview for Australian Life Scientists*
Dr Farah Zaib Khan, Scientific Business Analyst, Australian BioCommons

Public and freely accessible omics data repositories, such as those hosted by EMBL-EBI, NCBI, DDBJ, and others overseas, are a key part of the global life science research infrastructure. It is common practice for researchers around the world to access and reuse this data for scientific research.

Consultations undertaken by the Australian BioCommons over the past few years have identified that some Australian life scientists face challenges when attempting to access and download data housed in these offshore omics data repositories, particularly when this data is required at scale, or quickly. This presentation delves into these challenges and considers the potential of a national strategy to enhance the current situation regarding access to the data stored in these repositories.

About the presenter: Dr. Farah Zaib Khan, a Scientific Business Analyst at Australian BioCommons, is deeply committed to identifying and addressing the digital infrastructure needs of life science researchers. With her expertise in bioinformatics and IT, she adeptly facilitates communication between technical and non-technical stakeholders. At present, Farah is concentrating her efforts on collaborating with stakeholders to tackle challenges life scientists encounter concerning data submission to and download from international repositories.

11:20am | *Automatically acquiring instrument data & relocating repositories*
Craig Hamilton, Head of Technology, Intersect

Intersect has gained significant experience in data movement, research data management and data storage in supporting university members and research organisations. Research instruments are expensive and are often in high demand for use by researchers. How do you get the data off them in a reliable and timely manner, so that the turnaround time between projects is reduced as well as the data being made available to researchers for analysis? Craig will briefly review two use cases where we have implemented a tool to address this challenge.

Research data sets continue to grow, placing ever increasing demand on data storage capacity, Over the last several years, Intersect has moved significant volumes of data around the country, working with universities and research organisations to meet their storage needs. Craig will talk about how we did this in the past, and how we are doing it now, and show the impact of our learnings and tool choice.

About the presenter: Craig has 30 years' experience in Consulting, Engineering, Architecture and Product Management, across private, government and Higher Education organisations. From architecting and leading the engineering team that built the number one online retail shop of its time in Australia, to rolling it out to two other countries, Craig has proven experience in designing, building and operating scalable, secure systems. As an Enterprise Architect for multinational companies, Craig developed global strategies around Identity Management, cross-border ordering, and multi-faceted Customer Relationship Management. At Macquarie University Craig also designed their first Enterprise Service Bus, reducing integration between complex systems from weeks to minutes, and built their primary Data Hub, with integration and Master Data Management across a number of core subjects. Since joining Intersect in June 2016, Craig has been helping researchers manage and share their data, projects and knowledge, working with tools such as OpenStack, DMF, Mediaflux, AAF and software technologies, such as Java, PHP, Python, R, Ruby & Matlab.

11:30am | *Managing big data from a fleet of mobile laboratories*
Marcus John, Data Analyst, National Transport Research Organisation (NTRO)

Every government's most expensive asset is their road network. We are building a new way to collect and store terabytes worth of network level data on road condition for easy insights. In order to achieve this task we have integrated technologies from 5G to AWS, Starlink and Globus.

About the presenter: Marcus John has a Master of Science (Physics) from the University of Melbourne and is now working as a Senior Data Analyst and Software Developer at the National Transport Research Organisation (NTRO).

11:40am | *Panel discussion – addressing growing research demands*
Moderator: Gareth Williams. Jake Carroll, Dr David Poger, Dr Rebecca Farrington, Dr Farah Zaib Khan

As data volumes and velocity increase, are we prepared to manage the growing requirements of data-intensive research?

12:20pm | *Wrapping up – collaboration for evolving data challenges*
Greg Darcy, Digital Research Product Manager, AARNet

Is increased collaboration the solution to the evolving data challenges faced by facilities operating new scientific instruments and the researchers using them?