

# *ARDC Global experiences digital research infrastructure federations site visits - CERN, 2nd August 2019*

## Introduction

Monash University received Australian Research Data Commons (ARDC) infrastructure funding to lead a global site visit tour themed "Global experiences digital research infrastructure federations". We wish to thank CERN for participating as a site of this tour.

Confirmed members of touring panel includes Steve Quenette of Monash University (panel chair, and director of a university node of the Nectar Research Cloud), Max Wilkinson of the ARDC and Brendan Davey of TPAC (a director of a node of the Nectar Research Cloud). The panel has attracted collaboration of two other ARDC discovery projects. This includes Wojtek Goscinski, who leads the data focussed HPC facility (MASSIVE) and Komathy Padmanabhan, both of whom are delivering the "Machine learning infrastructure deployed at scale: understanding requirements, demand, impact and international best practice" ARDC discovery project. Additionally Anitha Kannan is joining us from the "Developing a proposal for enabling the creation of health data collections and a federated secure platform for research collaborations over these collections of sensitive data" ARDC discovery project. Anitha and Komathy will attend a subset of the sites. A member of the OpenStack Scientific Interest Group may yet still join the panel. These collaborations have enabled us to gain a broader scope of tour sites, and to better relate underpinning federated infrastructure to emerging research domains. We have provided bios below.

The premise of the tour is that our national OpenStack federation (the Nectar / Australian Research Cloud) has been an outstanding success and a global innovator. The ARDC is asking - what are the next major investments and what of the existing capabilities should sustained - to create a sustainable and impactful research environment. The intent of the tour is to amass international evidence that Nectar Research Cloud is still valuable, but also to amass international evidence that digital infrastructure federations either on-top of or along-side the Research Cloud for discipline-specific initiatives is also (a) proven, (b) prudent, and (c) value for money.

Over a 16 day period we will visit 13+ confirmed sites/infrastructures/networks, across 7 cities in the UK, EU and the USA. We seek to gain some insight into the tensions and implementation rationale of ...

- Federate at the bottom (a network of legally distinct infrastructures operating the same underpinning cloud infrastructure) vs federate at the top (a network of legally distinct entities operating the same domain-specific platform over potentially different clouds/infrastructure)

... which is intertwined with ...

- Infrastructure-led vs discipline-led
- Institution-led vs centrally-led
- Simulation- vs data- science
- Data gravity- vs privacy
- Private- vs public- cloud

The panel seeks to ensure ARDC have broad-stroke evidence of the environment necessary for many/all disciplines, knowing that in the Australian context, each discipline will dictate their own priorities.

Given the shortness of time and resources, we have targeted a few dominant domains such as: bioscience, imaging, physics and health / medical sciences. We have also targeted cross-discipline / tech-led technologies such as Kubernetes, AI, safe havens and OpenStack.

## Site visit tour questions

Our funding proposal explicitly states that we will ask each site:

### **DAP/T5/P10: What are the global experiences of infrastructure federations and what is the capacity of federation to continue to provide benefit?**

Questions to be answered include:

- Costs/Benefits of federation
- International models for federation – of data and/or tools and/or infrastructure
- Allocation and priority schemes – compare and contrast locally and internationally
- Governance models – compare and contrast locally and internationally
- How best to trade off between federation-wide standards for interoperability, and local requirements for customised services and policies?

In addition to the explicit DAP/T5/P10 questions, we committed to asking each participant:

- Was / is the Australian / Nectar Research Cloud noted as a reference model?
- Describe actual financial / co-investment / business model for your infrastructure / initiative?
- Describe end-user access / allocation policies for your infrastructure / initiative?
- How is federation technically and / or conceptually achieved in the service model for your site / initiative?

However we anticipate we will refine the questions as the tour progresses, and we are keen to work with sites on questions that are of interest to them. An example auxiliary question is:

- Are you interested in the tour outputs? If so in what form / forum?
- Does our / how can our federation interact with yours?

## DAP/T5/P8: What is the demand for machine learning tools and services and how could these be most effectively supported?

Over the past year, we have seen a steep increase in applied deep learning (DL) across our research platforms (MASSIVE and UQ RCC). Our proposal is to characterise the soft and hard infrastructure required to understand how to underpin and support this increasing adoption, at scale. Our goals are:

- Form a clear understanding of the relationship between research requirements, computing capability, capacity, and research impact. We will address this relationship across individual researchers, CoEs, infrastructure facilities (University facilities and NCRIS), and institution wide.
- Understand overall requirements, and how these can be consolidated across a cohort of researchers, and deployed in the most efficient and effective manner, at scale.
- Understand international best practices, and how it should inform Australian investment.

[More information.](#)

## Tour sites

At the time of writing - confirmed sites are:

- SAIL / Swansea University
- IRIS
- EOSC-Pilot
- Elixir
- Swiss Personalized Health Network (SPHN)
- Cambridge University
- Science and Engineering South consortium (SES)
- CERN
- SHRINE
- Center for Clinical Data Science
- University of Michigan
- Penn State University
- JetStream
- Indiana University's AI-HPC

## Panel members

Name and Contact	Representing
Global experiences digital research infrastructure federations panel members	
Dr Steve Quenette	<b>Monash node of the Research Cloud &amp; RDS</b>

<p><a href="mailto:steve.guenette@monash.edu">steve.guenette@monash.edu</a> +61 438 558 275</p>	<p>Dr Steve Quenette is the Deputy Director of the Monash eResearch Centre, and the lead of the Monash node of the Australian (Nectar) Research Cloud &amp; Research Data Services. His role within Monash University is to lead and drive the adoption of digital technology and advanced techniques in all fields of research. He has vast experience in business strategy, technology strategy, transformation and operations, with experiences ranging from universities, defence, hospitals and startups. He has a Phd in computer science, a Bch Science and a Bch Engineering. He has also participated in several applied domains of research including the biological sciences, healthcare, geophysics and energy. Steve has developed his present team to be global pioneers of software defined infrastructure, ranging from networks to clouds to domain-specific to federated platforms. He has long-standing partnerships with technology leaders and integrators.</p>
<p>Dr J Max Wilkinson <a href="mailto:max.wilkinson@ardc.edu.au">max.wilkinson@ardc.edu.au</a> +64 27 747 4890</p>	<p><b>ARDC Representation</b></p> <p>Dr Wilkinson is an independent consultant providing services to the research data management, research data governance and research infrastructure domains. For the last 5 years he has provided services to numerous eResearch organisations in Australasia including the National eScience Infrastructure (NeSI), Council of New Zealand Research Librarians (CONZUL), AgResearch, eResearch 2020, MBIE, the Australian Research Data Commons (ARDC) and Microscopy Australia. Prior to this he worked in the UK, 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 the NCRI. He received his PhD in Molecular Nephrology from UCL in 2003 and is a PRINCE2 certified project manager.</p>
<p>Brendan Davey <a href="mailto:brendan.davey@utas.edu.au">brendan.davey@utas.edu.au</a> +61 412 917 705</p>	<p><b>University of Tasmania node of the Research Cloud &amp; RDS</b></p> <p>Brendan Davey is the Acting Director of the Tasmanian Partnership for Advanced Computing (TPAC). TPAC provides eResearch services to local and interstate partners, and is a node of both the NeCTAR research cloud and Research Data Services. Brendan oversees the day to day delivery of eResearch services ensuring current and future eResearch needs are met. He has extensive experience in delivering ICT services in higher education and delivering multisite global ICT solutions in the commercial sector. He has been working in the eResearch sector for the past seven years and has managed many projects, many for the marine research community including the Marine Virtual Laboratory (MARVL), and numerous Australian National Data Services (ANDS) projects. He has a background in computer science and electronics. He is certified in ISO 27001 Information Security, BS25999 Business Continuity Management, Prince 2 and ITIL.</p>
<p>Machine learning infrastructure deployed at scale: understanding requirements, demand, impact and international best practice panel members</p>	
<p>Dr Wojtek Goscinski <a href="mailto:wojtek.goscinski@monash.edu">wojtek.goscinski@monash.edu</a> +61 412 133 156</p>	<p><b>MASSIVE</b></p> <p>Dr Wojtek James Goscinski is Associate Director at the Monash eResearch Centre, a role in which he leads teams to develop and implement digital strategies to nurture and underpin high-impact research. He is the coordinator of MASSIVE (<a href="http://www.massive.org.au">www.massive.org.au</a>), a national high performance computing facility for data analysis, and the manager of the Characterisation Virtual Laboratory (<a href="http://www.cvl.org.au">www.cvl.org.au</a>), an Australia program to connect scientific instruments with computing infrastructure. He holds a PhD in Computer Science, a Bachelor of Design and a Bachelor of Computer Science and Software Engineering.</p>
<p>Komathy Padmanabhan</p>	<p><b>Data Science and AI platform manager at Monash University</b></p>

<p>[USA part of the tour only] +61 424 711 587</p>	<p>Komathy Padmanabhan is the Data Science and AI platform manager at Monash University. In her role as the Establishment Manager, Komathy is responsible for developing foundational Strategy, Business and Operations model of the Data Science and AI platform. She has vast experience in Strategic planning, business architecture, Business-IT strategic alignment and Business performance management across multiple domains. Komathy holds Masters of Business Administration (in Strategic Management) and Bachelors in Information Technology.</p>
<p>Developing a proposal for enabling the creation of health data collections and a federated secure platform for research collaborations over these collections of sensitive data panel members</p>	
<p>Anitha Kannan [UK part of tour only] +61 409 926 465</p>	<p>Anitha Kannan is the Director, Research Platform Data Strategy in the Office of the Vice-Provost, Research and Research Infrastructure. Her role is focused on establishing and delivering a comprehensive data strategy for the Monash Technology Research Platforms in order to realise the full potential of the University's research infrastructure. This strategy encompasses technologies, processes and capabilities developed within the University and through strategic partnerships with national/international institutions, industry and government. She was instrumental in establishing Helix, which is a core capability supporting health researchers, the largest research cohort at the University, with health data infrastructure and informatics expertise. Prior to this, she was a part of the leadership team at the Monash eResearch Centre. This followed a 14 year stint in the IT industry with product development and IT services organisations like Hewlett Packard and Texas Instruments. She holds a Bachelor of Engineering in Electronics and Communication.</p>
<p>OpenStack Scientific Interest Group panel members</p>	
<p>John Taylor [some UK sites only]</p>	<p>John has been involved with HPC for over 30 years and witnessed the move away from proprietary Supercomputers to today's HPC commodity clusters. Most recently he co-founded StackHPC Ltd. to promote the convergence of HPC with cloud through the use and development of OpenStack within Research Computing. Previous to this John has worked in technical marketing, sales and project management for various HPC system and technology component suppliers. While at StackHPC, John has also been part of the Square Kilometre Array Science Data Processor management and design teams, under the aegis of the University of Cambridge. For the SKA, he worked on systems engineering artefacts and architecture definition and jointly led prototyping activities for large scale storage data challenges.</p>

# Agenda

[ADDRESS]

Hosted by Dr Tim Bell

Friday 2nd August, 2019

9:45 am - Arrival

10:00 am

- History of the Nectar Research Cloud [MW + SQ]
- Creating specialist facilities [SQ + WG]
- Outline of panel questions

11:45 am - Tea break

11:00 am

- CERN infrastructure
- Tour

12:00 pm - Lunch

1:00 pm

- ARDC funding landscape [MW]
- Panel Q & A and recap

2:00 pm - Panel depart