

Dr Phil Gurney  
Interim CEO, AeRO  
1-October-2021

To: researchinfrastructure@dese.gov.au

## Re: AeRO submission to the National Research Infrastructure Roadmap

AeRO, the Australasian eResearch Organisations, is the peak body for data-centric research across Australian and New Zealand. This document is provided as input into the road mapping process – to provide feedback on Australia’s digital research infrastructure, the gaps and pathways forward. It consists of a brief summary from the discussion at the AeRO Forum 2021, “Digital Research Infrastructure, Identifying the gaps”.

The Forum consisted of two panel discussions from key leaders in the sector, and attracted over 100 registered attendees who contributed to the discussion through the Q&A. The panel members were:

- Rosie Hicks, CEO ARDC
- Carina Kemp, Director, eResearch, AARnet
- Ray Fleming, Google Higher Education Strategic Projects and Program Manager, APAC
- Mark Stickells, Executive Director, Pawsey Supercomputing Research Centre
- Steve McEachern, Director, Australian Data Archive
- Andrew Lonie, Director, Australian BioCommons
- Merran Smith, Chief Executive, PHRN
- Grainne Moran, Pro-Vice-Chancellor (Research Infrastructure), UNSW

### Reflections on how things are working today in the digital research infrastructure space

There was a very positive energy from all the panellists and a key takeaway is that we have made strong progress in our journey to support researchers with appropriate digital infrastructures.

The digital onion, we were reminded, has - and needs - many layers.

Over the last funding cycle, many of our existing connections and collaborations between platforms, between researchers, and between institutions, have grown nationally and also internationally.

We have many jealous observers of our current capability and we are perhaps also victims of our current momentum with researchers hungry to use the next generation of data and computational infrastructures.

The community now recognises that the deluge of data growth is here to stay. While this brings challenges, the focus is switching to themes such as how we use infrastructure to support and improve **data quality, data re-use, and sharing and the value of data as a translational product.**

The community is adapting to ensure we can use data effectively at scale. **Data has gravity at the peak scale but data also wants to move** and so the challenge is to think through new Researcher workflows and break down data barriers which may take place across our platforms. We have great networks and need to use them effectively and get data moving intelligently.

The panel noted that the actual science methods are important – **we want to get good quality data**, not just data, and **metadata is of growing in importance** as a glue to bind. These are areas that still need work.

We also saw **data challenges around IP and jurisdictional issues**. Institutions and companies are putting up firewalls around data at the same time we are looking to use that data more effectively. But at the same time industry changes are making data products and data exchanges even more possible. **There is a genuine data economy internationally** and a challenge for our research platforms is to consider this new landscape while **being aware of the risks of Data Lockups by partners who increasingly see data as a cost and asset**.

Investments into our data infrastructures have provided quality outputs and a community that is much more aware of sustainability issues. In this context, **the sustainability of long data retention periods, required by legislations, is a strain on the sector**.

### Looking to the future

Over the last cycle our infrastructure, computational and discipline toolkits have improved. We have given our researchers access to new capabilities - **we still have some key capability gaps, especially in HASS and other areas where data-centric research is still relatively new**, but our landscape is more populated with higher quality tools than ever before.

There is an obvious **new set of capabilities now available through commercial computing as-a-service providers**. But **there remains a need to offer cost-effective national infrastructure**.

Taking advantage of the rapidly evolving commercial clouds is being considered by most of our disciplines, platforms and institutions. However we noted that **there are many devils in the detail of commercial infrastructure, including issues of OpEx budgets, commercial costs, data sovereignty responsiveness to research needs**. We also noted that **there is a skills gap relating to how best to take advantage of commercial clouds for research purposes**, and that this will need to be addressed if we are to gain the most out of what is on offer.

In addition, models **like providing funding direct to researchers to consume as-a-service providers are happening internationally**. We need to consider if these are appropriate in an Australian context, particularly as researchers seek to be involved in international collaborations. Our services providers are all looking at their sustainability and co- investment plans.

In regard to leveraging international expertise, it is clear that **while Australia has significant scale nationally, we are not an island**. In some areas, our scale is small and we need to look to those who have already reached scale – and **we need to learn not just those in our own discipline, but what is happening elsewhere, and happening internationally**.

Australia **needs to be wary about the cost of doing too many unique things**, particularly where we are seeking international collaboration. That said, we do lead in some areas and **many of our current tools provide great value to our researchers**. We need to keep the **focus on how innovation can add value while continuing to leverage large international discipline-based capabilities**.

One insight to consider is how researchers and facilities and emerging technology providers can **leverage co-design practices to solve emerging research problems**.

With **new cyber and institutional firewalls going up**, the challenges will be **how can our infrastructures support new research workloads without getting into reputational risk and protecting our IP and research outputs**.

Across all areas, **data governance, while boring, will be important for all of us.**

The panel noted that issues of scale remain, but these are now less about the scale of data or scale of compute (although these challenges are not insignificant!). The focus now is on **the scale of the application and toolkit landscape, and how to support a rapidly growing number of users.**

The next phase of investments needs to consider that we have scaled up largely as a result of helping solve the high-end needs of the nerds of the last generation. **We now need to support digitally native Early Career Researchers**, for example wanting to use ML for performance-arts VR projects. **Our toolkits are getting a lot more use.** As one of the panel members described the phenomena succinctly, we are now bringing haute-tech-couture to the high street.

In this context we noted that our **workforce is willing but small. Keeping and valuing this world class support expertise was a concern** of all panellists. We have great people supporting the infrastructure, but **too many research support staff have insecure jobs, and no career pathways**, jumping from job to job when programs disappear.

We also noted that **the focus for funding is increasingly – and in the most part appropriately – on the translation of research and the impact that our platforms deliver.** This is especially relevant to industry and government as consumers of some of the mature research products.

This is not changing the focus on basic research but all of our organisations are responding to the real landscape of how to pay for that basic research.

Given this context, **there is probably a concern here for some disciplines, especially those in HASS, who will be maturing their platforms and toolkits in the next cycle of funding as pencils get sharper.**

**Innovation and incremental improvement were important themes** as we talked about how to take advantage of new capabilities while maintaining support for the growing base-load that has evolved over the last cycle of investment.

We were very careful to note that **innovation in one area may be re-inventing a wheel that exists elsewhere** and given our new baseline we need to be careful to apply the lessons and tools we already have and to keep our tools focused – **AARNet is not Telstra; NeCTAR is not AWS, and all are set to play important roles in the future landscape.**

In this context the N in National infrastructure was highlighted and **the benefits of National toolkits and National scale should be retained** and even more leveraged going forward.

At the same time, there was a recognition of agility and the ability to target and quickly re-target our national investments and resources was also seen to be important. **Building big things takes time and the stop- start nature of funding is challenging.** Sustained funding plans enable our capabilities to target quickly emerging issues such as CoVid or natural disasters.

Lastly, we reminded ourselves that the C in NCRIS is about collaboration – we need to build to allow researchers to collaborate locally, nationally, and internationally, noting that other jurisdictions will inevitably take a different path to us.