



AeRO Forum 2015 – 22 July, University of Canberra

Afternoon discussions summary

The AeRO National Forum provides a venue bringing together a wide range of stakeholders and other interested parties across the eResearch sector. Around 85 attendees participated this time and included national and state eResearch service providers, NCRIS capabilities, funding bodies the ARC and NH&MRC, national bodies including CSIRO, GA, ANSTO, NLA, DSTO, overseas initiatives such as NeSI(NZ) and NZ-Genomics, bodies including CAUDIT, Universities Australia (UA), Science and Technology Australia (STA), the Australian Academy of the Humanities, the Australian Academy of Technological Sciences and Engineering, and many institutional representatives from eResearch and the wider IT services groups.

This Forum looked at the national Research Infrastructure Review (Dr Susan Pond), the recent report on current eResearch initiatives (Prof Tom Cochrane), and an indication of future analyses of eResearch service requirements and guidance (Dr Rhys Francis). These presentations led to two discussion sessions in the afternoon, which are summarised below.

Discussion #1 - Direction Setting

The first discussion topic was to identify the issues that we need to prepare ourselves for, collectively, for evolving research requirements and needs in the context of rapidly evolving e-infrastructure technologies. Questions leading the discussions included:

- What has changed in research over the decade since NCRIS started? What has remained?
- What trends (research and technical) could have a major impact over the next 5-10 years?
- What would need to be in a decadal plan for eResearch?

The discussions roamed widely and freely, with some points revisited at various time. The following summary structures the comments under various headings.

Context

Many comments revolved around two particular broad themes: 'Data' and 'Scale'. The central role of 'data' was obvious to all, however the issues and opportunities differed depending on the context for individual contributors. Comments included:

- Compute (analysis) is increasingly moving to the data. This includes overseas facilities in some cases.
- We need to work out the best ways to access data, regardless of location, and how/if to bring compute to it. We need agreements in place to enable us to access it either way.
- Data re-use is increasing, and more people are recognising that data increases in value through circulation. Synthesis/aggregation of data from multiple sources is also increasing, which creates additional issues.
- There is increasing demand for openness around data, for a wide range of reasons – policy, public good, etc.



Scaling (around data services) was another common theme.

- Data volumes are increasing rapidly through multiple mechanisms, including:
 - Modelling creating more data
 - Sensors becoming bigger, higher-resolution
 - Sensors becoming more numerous, especially with the Internet of Things (IoT)
- This will increase the need for automation
- It will increase the need for making it seamless
- Scaling is not restricted to data but also affects user numbers as they are increasingly 'pushed'/drawn into eResearch.
- Linking these discussions was the observation that use of "Big Data" is enabled by tools, rather than by the data.

Responsibilities/Roles of research institutions

With a large number of investments into e-infrastructure, across multiple providers, a theme that emerged was the role of institutions. Attendees noted that it was not clear that anybody could articulate 'responsibility' for eResearch capabilities, at any level, however institutions clearly had a crucial role. To support them meant understanding their issues:

- **Expectations** – there is no clear/common view regarding the expectations of institutions in relation to what research e-infrastructure they support, and how, and there will be a diversity of views.
- **Competition** – we cannot stop institutions from competing with each other through their investments. However it would be helpful for the institutions, the sector and providers to understand the value returned.
- **Internal funding** – Universities already feel that they "bleed" other areas of funding in order to fund research.
- **Investment** – some institutions are investing heavily, compared to others, with some investing in local e-infrastructure, others more in frameworks, e.g. policies or end-to-end data management support, working hands-on with researchers. In some cases the development is nuanced, in eResearch services rather than eResearch infrastructure.
- **Constraints** – it was noted that researchers using desktops inside a campus are both directly supported and sometimes constrained by the campus infrastructure. This can hamper their access to national infrastructure.
- **Funding cycles** – a challenge for institutions is that institutional funding cycles and plans are typically 1-2 years only, not recognising that this creates problems for planning, for investing in long-term services and for supporting a specialised workforce.



Workforce

A very strong theme emerged regarding workforce development, building skills in both the service providers and in the user (research) community. There is a clear perception that there is a shortfall in the capacity of the sector to support users.

- Finding skilled people is a major problem.
 - Universities do not appear to be creating the types and number of people required to provide support in this sector.
 - Researchers are (generally) seen not to be picking up the skills, or quickly enough, to take advantage of all the opportunities, when it was appropriate.
- We do not have, but need, a framework for skills/workforce development.
- An important issue for developing support staff was there appear to be no visible career paths in this space.

Funding, investing

Given the significant investments in this space, a discussion around funding and pricing highlighted the following important points:

- **Costing models** – the distinction between ‘capital’ and ‘operating’ costs, as set out in funding rules for some projects, was extremely unhelpful, and in the case of e-infrastructure/services sometimes almost a meaningless distinction.
- **Business models** – the e-infrastructure that has been created needs to have associated business models that support ongoing operations and long-term access promises. Some attendees felt that some infrastructure has been built without establishing prices, which effectively creates subsidies (“distortions”) in areas where we may compete, e.g. against market offerings.
- **Metrics** – other attendees noted it is difficult to fully associate prices/costs with all of the benefits, which is what the funders also measure through a range of metrics. We don’t spend money on research to save money, but to enable/accelerate research and achieve impact. Reducing time to reach an outcome is one benefit.
- **Importance of scale** - We don't scale if our methods require us to sit down 1:1 with each researcher; the costs will be too high, putting pressure on funding.



Targets/Scope for investment

A large part of the discussion revolved around the best ways to invest for any particular target audience. The role of the commercial sector in providing e-infrastructure services was canvassed.

Market offerings:

- **Value-add** – it was suggested that there could be more focus on “real value-add” or more specialised services, and let commodity take care of what it does best, however there was a significant diversity of views on that point.
- **Leveraging offerings** – some providers were considering what was being offered in the market and how that could support them. It was noted that NCI provide highly integrated environments but could move elements to commercial services if the support was there.
- **Gaps** – many felt there were too many gaps still in market offerings, especially around integration, so our sector may need to lead the market to new opportunities
- **Trust** – one discussion noted that Medical/Health data needs trust. It was not clear whether researchers could contract (at all) for the necessary levels of trust, or if the prices were then unreasonable.

Market models, build or buy

- Some noted that there is always a build-or-buy decision, always has been and always will be.
- Some commercial entities (as clients) are building their own private infrastructure (clouds for compute/storage), but they also partner with public cloud providers for burst capacity.
- It was pointed out that over the last decade the sector has built many new tools and then been overtaken by industry, repeatedly. This may suggest that fast-following is a potential model, if researchers’ requirements can be met.

Who are we building for?

Attendees discussed potential investment models:

- Investment in deeply integrated facilities
- Investment along technology lines, or
- Investment targeting specific communities.

A diversity of views suggested all were appropriate, depending on the context. Some major discipline areas, such as some of the national research priorities appeared to be well supported by e-infrastructure integrated specifically for the needs of that community. Others noted this suited some ‘peak’ communities, however other communities (e.g. HASS) were yet to fully engage with the e-infrastructure opportunities and had needs that spanned facilities, technologies and locations. This reinforced the earlier scaling discussions, where providing for individuals and small-scale research groups is a larger challenge.



Planning

The current reviews and framework development activities highlight the importance of planning and coordination, and that is a multi-party conversation in our sector. It was noted that the international NREN (National Research and Education Network) community were striving towards a unified global architecture and had made progress. A clear trend is that multidisciplinary research is increasing rapidly, as evidenced in major publications.

- **Strategic planning** – developing a more strategic plan for e-infrastructure was seen as challenging, but nonetheless feasible. It must be flexible, prepared to change, and not too specific. It could set out the aspirations of the community - the main things we want to achieve over the next ten years. It should allow for a diversity of funding mechanisms, service providers (including commercial), and both build or buy options as appropriate at the time.
- **Priorities** – our sector is challenged in deciding what to start as well as what to stop doing. This probably requires broader input, including about the (research) impact of the (e-infrastructure) investments.
- **Planning and governance** – it is not clear whose role it is to support planning, and what kind of “governance” this sector needs for any improved organisation, communication and collaboration. It was suggested that AeRO can provide advice and/or some leadership in this space; this was left for the subsequent discussion topic.
- **Gaps** – there are many gaps (“whitespace”) in the national e-infrastructure that need to be mapped and need to be addressed. One example noted the increasing demand for open data within a broader open-science context, and that both data and workflows become part of the publication. It is not clear who can do and support that, since it involves multiple service providers.

Innovation

Another theme raised the importance of innovation within the e-infrastructure sector. The government refers to a Research and Innovation System, while most of our community focusses on research. It was suggested that we explore opportunities to connect more into the Innovation system, where there may be more opportunities for investment into e-infrastructure. Several large companies (e.g. IBM, Cisco, Bosch) are heavily involved in connected devices, sensors and the broader Internet of Things (IoT). They require significant e-infrastructure capabilities to both support and leverage that.

Software/information companies are increasingly dominating other sectors in the market. All of these are relevant for researchers, and could be leveraged for the benefit of researchers and the benefit of the commercial sector. It was noted that some universities have established incubator centres, and there are other examples of e-infrastructure engaging strategically with industry, creating additional revenue streams.

Another area growing from the government sector is the increasing visibility of state and federal data.X.gov.au catalogues, various hackathon events and on occasion even startup companies arising from these datasets and tools.



Discussion #2 - Integrating across boundaries

The second discussion topic looked at issues where e-infrastructure presented boundaries to users and providers, which in essence create friction for researchers trying to use them. The infrastructure and services are provided at the institutional level, through multi-institutional partnerships, and through national and international providers. Again the conversation ranged widely and has been aggregated and summarised below.

Context

It was noted that we needed clarification of the types of boundaries we faced.

- Even integrating within boundaries (e.g. within an institution) is still hard, and the boundaries change with time. In the experience of the AAF they support about 250 services (40% of them for research) that cross institutional boundaries, and they see an ever-evolving landscape on both the technology and the relationship boundaries they need to manage.
- We probably should be talking about integration within a project, within a community, or both, and there are many levels within either approach.
- Pragmatically, integration is hard as everyone is busy, contracts are short, and so integration can become a distraction.
- Data policies are a serious boundary, limiting data access and re-use across projects, communities and sites. Some government agencies hold data of interest on their own systems and are heavily constrained by policy from providing in-situ access, let alone letting data move to other facilities for re-use.

Being strategic

Considering the multiple investments and pathways that have occurred, three key elements were suggested to facilitate better integration:

1. **Communication** – improved communication at the beginning of projects. JISC was noted for bringing common interests together at the beginning, to ensure awareness, collaboration and coordination were streamlined.
2. **Standards** – the adoption of standards, which are ‘owned’ and ‘promoted’ by someone in a position of some authority. ANDS was noted as an exemplar with their development of RIF-CS, and that it leveraged the appetite for data exchange to great effect.
3. **Software** – a more strategic approach towards software is required – especially regarding software re-use. Different funding bodies have led to similar software being developed multiple times, in different places.



Better governance/oversight

The discussion touched on aspects of governance and oversight, and noted that this applied both to the various projects as well as the institutional frameworks they connect with.

- **Lead agents** – the lead agent model perhaps did not ‘include the essence’ of collaboration from the beginning. It could include a requirement to specify how they are going to collaborate and coordinate, across and within projects.
- **Governance** – we do not have standardised governance models across the various investments, nor been provided guidance on how things should operate.
- **Leadership** – it was highlighted the NZ initiatives had independent directors and government observers on the board, with a clear understanding that this is an investment by the government, rather than a contribution. Where cross-program initiatives were required they established a joint working party, with senior (CEO/Board level) representatives from all sides. This sits alongside their governance structures, not above nor below.
- **Broader forums** – Canada recently established the Leadership Council for Digital Infrastructure which brings together parties from institutions, service providers, communities and funding bodies. This provides a forum for discussion around integration and policy development.
- **Engagement** – Institutions have their own challenges here. Policy sits with the DVC-R and the CIO and the Librarian and the Research Office, and possibly more. This creates an engagement problem within institutions and across them through their various ‘representative’ bodies (CAUL, CAUDIT, ARMS, ...). There is an opportunity for strengthening the research management function within universities.
- **Management** – for individual activities, the use of more formal project management approaches and more agile methods (software engineering) could provide mechanisms to engage with broader coordination.

Engaging stakeholders

A common point was that stakeholders, in particular users/researchers, help to define the needs and expectations for integration. This creates a need for those users to have sufficient awareness of the opportunities.

- **Focus** – some universities (e.g. Melbourne) have activities especially focussed on early career researchers/HDR students – or more broadly the “young or the young-at-heart”. Others work to upskill the technical support staff. This creates scaling problems.
- **Training** – a few noted that they were considering (or even deploying) train-the-trainer models, and this related back to the broader workforce development discussion earlier. Materials being shared can be extremely helpful.
- **Champions** – it was pointed out that to achieve change you need to identify the champions and change agents to work with them, and also seek the support of the leadership to allow this to be developed. This was often hampered by limited institutional funding support.



Coordination

When the original Platforms-for-Collaboration NCRIS capability was defined, it was decided not to establish a 'head office' at the time due to uncertainty about the future. The eResearch-framework project will develop greater certainty in some respects, and it was noted other NCRIS capabilities have a 'head-office' or other form of over-arching guidance. This triggered a wide ranging discussion around the opportunities and challenges, comments included:

- Having a central entity could reduce the time it takes to get projects up and running.
- A body to assist with data policy coordination and alignment would help release data held in inaccessible agencies and centres.
- Driving execution with alignment needs something in the middle to deal with the gaps between functions.
- Maybe one does not need a head office to coordinate, but could perhaps allow the expectations of key stakeholders to drive behaviours, or through policy.
- Setting up any mechanism to drive integration will require additional funding, since it is not allocated within any current budgets. It will also need time to achieve things, and the mechanisms to make decisions about which gaps to tackle when.
- It was noted that any 'head office' or guiding framework had to include service providers, institutions (at multiple levels), government agencies, industry etc. to ensure full and proper communication.