eResearch AUSTRALASIA2019

An eResearch Strategy for a National Organisation

Carina Kemp

Carina Kemp[1], Ingrid Mason[2]

[1] AARNet Pty Ltd, Canberra, Australia, carina.kemp@aarnet.edu.au
[2] AARNet Pty Ltd, Canberra, Australia, Ingrid.mason@aarnet.edu.au

SUMMARY

This paper explains AARNet's eResearch activities and the strategy to understand the national digital research infrastructure requirements into the future for Australian Researchers. This eResearch Strategy will enable AARNet to map the research requirements against existing cloud infrastructure and service development at AARNet and plan for future development.

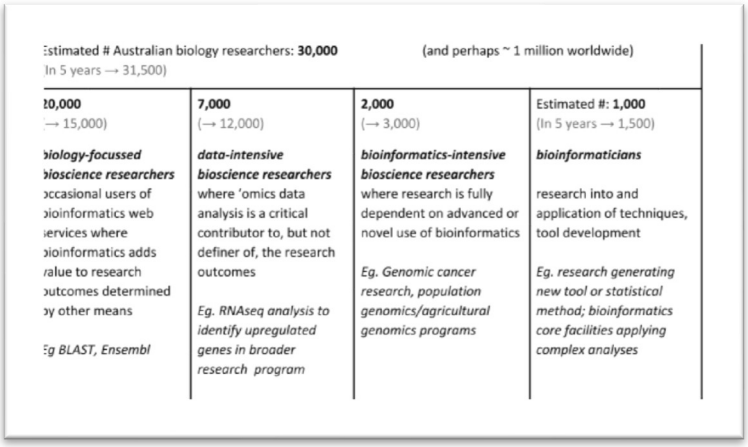
THE HYPOTHESIS – TRANSITIONS TO CLOUD FOR RESEARCH

EResearch activities at AARNet have been testing the hypothesis that despite the differences, in both scale and complexity, in the way that research across the different thematic domains is carried out there are fundamental infrastructure requirements for network bandwidth, data storage and compute that cut across all of Australia's major research domains. Cloud Infrastructure to deliver Storage and Compute services are now very mature technologies in Australia. This eResearch activity is also testing the hypothesis that "cloud first" for research is relevant to *all* research domains and *all* scales of data and technology usage, despite a large number of Australian researchers still using their own personal computers or small on premise[1] servers to do their research.

The uptake of CloudStor[2] by researchers (65,000 over 4 years) indicates that a common cloud platform (storage and compute) operates effectively as *cloud infrastructure enabled* research. In the next phase of work undertaken in eResearch at AARNet, the move is from hypothesis testing to evidence based decision-making. The aim being to understand how and when cloud storage and compute will become central to research, by examining the growth edges of clusters of research domains. It was discovered that four domain clusters examined operate across four generalised scales of cloud infrastructure usage and in parallel these scales approximately map to four scales of digital capability. This was illustrated by a study by the Australian Bioinformatics community in 2017-2018 where they mapped the Australian Biosciences Researchers across 4 scales of digital capability from occasional use of web based tools through to research and investigation into big data applications and tools development, see Table 1 below.

[1] Grant funding mechanisms support research groups buying their own capital hardware rather than using a central compute resource.
[2] AARNet's Cloud Storage and Compute Platform

Table 1: The digital capability of Australian Bioscience Researchers mapped in 2017-2018 and the estimated key transitions needed to lift the bioinformatics capability of the Domain. Source: An Australian Bioinformatics Commons Pathfinder Phase Project Outline, 2018.



Estimated # Australian biology researchers: **30,000** (and perhaps ~ 1 million worldwide) In 5 years → 31,500)			
20,000 → 15,000) ***biology-focussed bioscience researchers*** occasional users of bioinformatics web services where bioinformatics adds value to research outcomes determined by other means *Eg BLAST, Ensembl*	**7,000** (→ 12,000) ***data-intensive bioscience researchers*** where 'omics data analysis is a critical contributor to, but not definer of, the research outcomes *Eg. RNAseq analysis to identify upregulated genes in broader research program*	**2,000** (→ 3,000) ***bioinformatics-intensive bioscience researchers*** where research is fully dependent on advanced or novel use of bioinformatics *Eg. Genomic cancer research, population genomics/agricultural genomics programs*	Estimated #: **1,000** (In 5 years → 1,500) ***bioinformaticians*** research into and application of techniques, tool development *Eg. research generating new tool or statistical method; bioinformatics core facilities applying complex analyses*

DIGITAL RESEARCH TRANSFORMATION

Digital research transformation is occurring through the introduction of new technologies, infrastructure, collaboration, and skills. In Australia, this transformation process (referred to as "eResearch") over the last decade has been dominated by the federal government investments in facilities, large instruments, data generating equipment, and policy interventions. The reconfiguration of the NCRIS investments and 2016 Roadmap leans towards consolidation and a concerted shift into cloud infrastructure and services underpinning and driving change in research practices.

The rationale for "cloud first – for research" shares similar rationalisations with "cloud first – for enterprise" as part of the digital transformation of higher education. A shift away from institutionally hosted and maintained infrastructure and services, and efficiency gains made by outsourcing, moving to new technologies built to work with web technologies and operating on highly available cloud services. The bespoke workflows of research, distinguishes research requirements from enterprise wide requirements (in the uptake of cloud infrastructure and services). There are common workflows to be found across research domains, and gathering these in-common requirements is part of the process of determining what constitutes market failure (that informs a "build" strategy) and what can be best served through the uptake of commodity services (that informs a "buy" strategy).

The ultimate measures for this move into cloud are:

- All researchers and research domains gain from the infrastructural turn from cloud-enabled to cloud-centric infrastructure (research relevant and mass uptake of common infrastructure, enables wide collaboration and higher utility).
- Institutions are well informed to make an investment decision that can deliver to all of research (infrastructure designed for and of value to all research that leverages commodity offerings, enables cost-effectiveness).
- The research undertaken using cloud centric infrastructure raises the capacity for research impact (new techniques and innovation) and ensures Australian researchers are global competitors or partners (aligning infrastructural capacity and community capability with global trends, enables agility).

WHAT THIS MEANS FOR AARNET

AARNet has focused the design and extension of the NREN and the layered services to meet the needs of research that marry with specialised requirements for access to peak facilities or unique instruments and services (e.g. Science DMZ, CloudStor nodes at Peak facilities, sensitive data) and generalised requirements for cloud storage, compute and movement services (e.g. CloudStor and Jupyter notebooks). Of necessity, we are now moving from working with a rough grained understanding of requirements for cloud infrastructure and services, to one that needs to be more granular. This adjustment in approach asks more from the shareholders and the company in terms of global and national positioning, customer consultation, market analysis, and financial modelling. The AARNet model supports shareholder collaboration and co-investment to reduce marginal costs, rapidly scale, and work co-operatively to integrate enabling technologies for research. Through the domain engagement the vision for CloudStor as the connective and common infrastructure that has arisen from prior investment will be anchored to projections for cloud storage and compute across multiple domains. It is envisaged that this will enable efficiency across our member institutions (shareholders), accelerate existing research, and support structural change required for the digital transformation of research.

Cloud infrastructure and services, as layered services on the network, will become critical underpinning national research infrastructure over the next three years for all research domains.

REFERENCES

1. Lonie, A, et al., An Australian Bioinformatics Commons Pathfinder Phase Project Outline, per comms, 2018.