

Draft 2016 National Research Infrastructure Roadmap

SUBMISSION BY THE DATA RESOURCES WORKING GROUP OF THE ECOSYSTEM SCIENCE COUNCIL (ESC)

The draft roadmap correctly identifies digital data and eResearch platforms as an important focus area. Envisioning this research infrastructure along the lines of the European Open Science Cloud is a good model. However, the framing for Australia appears unduly narrow and misses the opportunity to innovate in terms of data reuse and attempts to meet the needs of all domains with a single approach. All of these can easily be lessened as proposed below.

Suggested improvements

1. Overall gaps in scope

The [Productivity Commission Draft Report](#) (final due in March 2017) has some key findings that are relevant to the NCRIS draft roadmap. To address the gaps identified below we suggest consulting the The Productivity Commission Draft Report.

- A mature capability would demonstrate its usefulness through uptake of research infrastructure. However, uptake is weak – for instance, around 22% of the digital data generated globally was potentially useful as an input into subsequent analysis but less than 5% of that data was actually analysed (EMC Corporation 2014).
- As a result of poor uptake new data management frameworks are required for more consumer oriented eResearch infrastructure. The rationale for this different approach is that incremental change to current data management frameworks will not suffice; instead fundamental and systematic changes are needed to Australia's data management frameworks.
- The framework needs to adapt to rapidly evolving data coverage, data standards , data content and data volume from a perspective that incentivises and supports reusers rather than for its own sake. Australia requires a framework for public and private datasets (and their curation) to be nominated and designated as National Interest Datasets (NIDs), and be protected by legislation.
- Australia requires a process for identification of high value datasets which are generally characterised by their uniqueness (e.g. cannot be replicated), high quality, or high degree of coverage in the relevant population, high frequency of updates. Such datasets will also be well described and understandable for intelligible reuse.



2. The Research Infrastructure National Advisory Group (RINAG)

- We support the establishment of the RINAG but would like to see further details about membership selection (nominations and community voting) as this would build confidence its independence.
- NCRIS has already enabled data to become more widely available and encourages the use of the FAIR principles (findability, accessibility, interoperability and reusability). However, in the report an increased focus on reusability of infrastructure for ALL stakeholders is critically important.
- The scope of RINAG should reflect the well-established research integrity and IP policies of existing entities, particularly in the science domain. For example, the ARC has established policies to do with research data (<http://www.arc.gov.au/policy>) as do external bodies (e.g. [World Data System](#)s guidelines) in addition to ANDS's reflection of Research Data Alliance's set.

3. Digital Data and eResearch Platform

- We endorse the establishment of the Australian Research Data Cloud infrastructure as a coordinated approach linked to HPC. We would prefer to see this capability facilitate integration of the hard infrastructure across the other eight focal areas (Table 2) rather than maintain each research domain pursue its own technical solutions. For example ACCESS, BCCVL and the ALA spatial tools overlap considerably and could be amalgamated into a set of more uniform national visualisation tools, which would benefit the end-user in terms of efficiencies in reuse. The data repositories, however, would be relatively independent and customised while sharing the hard infrastructure, technical solutions and services. A good example of success in delivering integrated solutions in the environmental domain is the UK NERC's [Ecology and Hydrology Centre](#).

4. Environmental Systems

- The emphasis on biodiversity data and the ALA platform is framing bias. There are platforms producing excellent data but most of all we need in situ long-term biota and environment data for policy, management and reporting about adaptations. The ALA platform business model is part of the solution but not complete for consumer-oriented infrastructure.

