# National Research Infrastructure Capability Issues Paper 2016 consultation

# Response by CAUL
# 9th September 2016

| Name | Harry Rolf |
|---|---|
| **Title/role** | **Communication and Policy Officer** |
| **Organisation** | **Council of Australian University Librarians (CAUL)** |

This submission was prepared with the assistance of CAUL members, and support from the CAUL office.

Prepared by Harry Rolf, Communication and Policy Officer, CAUL
Contact: (02) 6125 2990 or cpo@caul.edu.au

# Overview

CAUL welcomes the opportunity to contribute to the National Collaborative Research Infrastructure Roadmap 2016, and the shared vision of Australia's national research infrastructure. This submission addresses questions posed in the issues paper, and makes the following recommendations:

1. That Libraries constitute important national research infrastructure, and that this is recognised in the national research infrastructure roadmap.
2. That a plan be developed for exposing publicly funded research outputs and making them accessible to the world. The plan should consider:
   a. the creation of a distinct national discovery service for research outputs to link existing institutional repositories,
   b. the establishment of a national framework for the further development of research repositories, to promote their consistency.
3. That the Roadmap recognises Trove as critical research infrastructure.
4. That the Roadmap recognises the importance of services that provide researchers with the capability to 'publish' their data, and that enable access, management and standardisation of data publishing across Australia and internationally.
5. That the Roadmap supports the adoption of an appropriate model for building Australia's digitisation capability that considers the issues raised in this paper.
6. That the Roadmap recognises the importance of services that enable researchers to acquire skills and competencies essential for them to access and effectively use research infrastructure.
7. That the governance of national research infrastructure is a priority, including putting in place policies, standards and practices to:
   a. make publicly funded research outputs findable, accessible, interoperable and reusable (FAIR);
   b. value and practically support a range of dissemination models, recognising discipline diversity, whilst maintaining a commitment to the principles of this approach; and;
   c. ensure that policy is aligned with national codes of conduct and other codes of practice which lay out requirements to disseminate research responsibly, ethically and for the benefit of the Australian and international community.

# Introduction

The Council of Australian University Librarians (CAUL) is the peak leadership organisation for university libraries in Australia. CAUL seeks to enhance the value and capacity of Australian university libraries to influence scholarship, learning, and information policies and practices relevant to Australian higher education. CAUL members are the University Librarians or equivalent of the 39 institutions that have representation on Universities Australia.

University libraries are diverse institutions, which have important roles in many aspects of core university business. This includes responsibilities for managing, measuring and increasing many aspects of research such as outputs, researcher capabilities, and research practice. This makes libraries an important part Australia's research infrastructure nexus.

Australian university libraries are significant components of research infrastructure within their institutions, enabling research capabilities fundamental to the conduct of research. Nationally they form a network, serving researchers both within and external to their institutions.

A library acts as a hub connecting diverse human, digital and physical infrastructures, providing sophisticated services for research and innovation through:

- ❖ provision of repositories to store, promote and preserve the digital assets and outputs of universities;
- ❖ management, and increasingly creation, of research resources such as datasets and digital collections;
- ❖ publication, especially electronic publication, of material based on research;
- ❖ support for innovation in scholarly expression and communication;
- ❖ services for measuring and increasing research impact used in audit processes such as HERDC or the ERA;
- ❖ provision of advice on data management policy and planning, metadata, standards and persistent identifiers; and,
- ❖ contribution to the development of lifelong learning, research specific and transferable skills for researchers and students i.e. for using research infrastructure.

The 'library' is an important piece of research infrastructure – comprising physical and virtual places, collections and staff. The university library does not operate in isolation. University libraries are highly connected, collaborating with one another and with their stakeholders – both individuals and organisations – collectively providing important distributed research capabilities.

University libraries as national research infrastructure fits within the definition of research infrastructure used in the National Collaborative Research Infrastructure Strategy Strategic Roadmap 2011[1], where infrastructure comprises:

> *'the assets, facilities and services which support research across the innovation system and which maintain the capacity of researchers to undertake excellent research and deliver innovation outcomes' (p.5).*

---

[1] National Collaborative Research Infrastructure Strategy Strategic Roadmap 2011. Retrieved from https://docs.education.gov.au/node/34121

Today university libraries, as with other national infrastructure, find themselves operating in an environment subject to change and uncertainty, in some part due to the disruptive forces of new and pervasive digital technologies.

The following three parts of this submission address the topics of research capabilities, infrastructure, and governance more specifically. They address questions posed by the issues paper and provide comment on specific issues from the perspective of Australia's university libraries.

# National research infrastructure capabilities

This part of the submission addresses the following questions from the issues paper:

- ***Questions 1:*** *Are there other capability areas that should be considered?*
- ***Question 5:*** *Should research workforce skills be considered a research infrastructure issue?*
- ***Question 6:*** *How can national research infrastructure assist in training and skills development?*
- ***Question 7:*** *What responsibility should research institutions have in supporting the development of infrastructure ready researchers and technical specialists?*

The issues paper broadly and successfully captures areas of research where capabilities are required, and it is commendable that the paper includes a specific focus on 'understanding cultures and communities' addressing research capabilities for both HASS and STEM disciplines.

From the perspective of university libraries an important challenge and broad theme in the context of research capabilities and infrastructure is 'Information and digital capabilities' for research. The theme emerges from the challenge of pervasive and emergent digital technologies that are having far-reaching impacts on the research process. Technology trends such as those identified by the NMC Horizon Report: 2015 Library Edition have implications for both university libraries and researchers (Johnson, Adams Becker, Estrada & Freeman, 2015).

Bringing information and digital capabilities together offers a different way of viewing research capabilities. Four key areas of capability relevant to this theme are:

(1) the changing nature of scholarly records;
(2) research data management;
(3) digitisation, imaging and visualisation; and;
(4) skills and literacies.

Some of these capabilities are picked up under the 'data for research and discoverability' and 'underpinning research infrastructure' theme areas identified in the issues paper. However, here the importance of the four capabilities and relationships between them are emphasised and expanded upon.


## (1) The changing nature of scholarly records

The changing nature of the scholarly record requires effective stewardship for the capture, collection, description, discovery and preservation of the traditional print and digital content (whether it be format shifted or born digital). Expertise in the application of metadata, digital preservation standards, licensing, copyright as well as curation and archiving are essential to prevent loss of content through maladministration, degradation, physical damage or technological obsolescence. Given the importance of scholarly records and repositories to research it is timely for a plan to ensure that publicly funded research outputs are exposed and made accessible to the world.

Historically Australia has embraced institutional repositories through government funded projects such as the Australian Research Repositories Online to the World (ARROW)[2] project, the Australian Digital Theses (ADT)[3] program and the Australian Scheme for Higher Education Repositories (ASHER)[4]. These projects have provided support for the development of repositories in individual institutions. With these schemes now ceased, and no current national scheme in place, institutional repositories and practices may begin to vary in consistency and approach.

From 2009-2014 an annual survey of Australasian institutional research publication repositories has been carried out by CAUL, an initiative started by CAIRSS (CAUL Australasian Institutional Repository Support Service).  CAIRSS was active from 2009 to 2012 initially supported by ARROW funds, later supported by CAUL's redirecting its funding of the ADT program.

The 2014 survey shows that Research Publication Repositories (RPRs) are primarily (91%) consolidated within the university library. It shows RPR management moving away from combined or shared arrangements between the library and other departments such as the research office, and that at no time had any university's research office been solely responsible for the institution's RPR.

According to the 2014 report, usage of the RPR by institutions was also increasing, 'reflecting growing recognition of the utility of the repository as an institutional asset' (CAUL 2014, p.16). Usage metrics are used for academic promotion, budget purposes and internal and external reporting. Universities are now requiring a greater range of metrics, particularly the capability to track *Altmetrics* and new forms of research impact and engagement.

The recent removal of research publications from institutional HERDC reporting requirements[5] is likely to make institutional approaches to RPRs even less consistent. The CAUL survey of RPRs showed that nearly 80% of institutional repositories were involved with the ERA and more than 50% with HERDC data collection.

An opportunity exists for reviving the concept of a distinct national discovery service for research outputs. This could be achieved through a linked data approach leveraging the web to expose, share and connect this data similar to that taken by Trove, or even through Trove. Trove began as a project by the National Library of Australia to provide a portal to library, gallery and museum collections. Today Trove is a world class example of innovative research infrastructure, deserving recognition of its status in the national roadmap, and appropriate resourcing so that it can deliver additional benefits to the nation's researchers by further developing its capability.

Other capabilities also go beyond the capacity of university libraries. A move towards open access scholarly records is making the copyright environment increasingly complex as funding agencies and

---

[2] the Australian Research Repositories Online to the World: http://arrow.edu.au/about/

[3] Australian Digital Theses program:
http://www.caul.edu.au/caul-programs/australasian-digital-theses

[4] the Australian Scheme for Higher Education Repositories:
http://www.industry.gov.au/science/ResearchInfrastructure/Pages/ASHERandIAP.aspx

[5] HERDC reporting requirements: https://docs.education.gov.au/node/39431

publishers adopt open access principles. Services and tools have been widely adopted by institutions to manage open access compliance issues. In 2014, most institutions provided links through their RPR to SHERPA/RoMEO, a service that provides summary of "permissions that are normally given as part of each publisher's copyright transfer agreement [with researchers/authors]" on a journal-by-journal basis. With funder open access deposit mandates coming into play, other services such as SHERPA/FACT (Funders' and Authors' Compliance Tool) or SHERPA/Juliet (a database of funders' policies and open access requirements) is now also important.

These services allow researchers to quickly and reliably determine whether the journal they wish to publish in will support compliance with the mandates of relevant funding agencies. With moves afoot to put these services on a more sustainable financial basis, Australia will need to find a way to take full advantage of these services. By either buying in to existing approaches or finding another way to ensure that its funding agencies' and publishers' requirements are included in a single international database.

## (2) Research data management

CAUL commends the issues paper's emphasis on research data management. Appropriate management of high quality data is essential to meet compliance requirements and to make the outputs of publicly funded research open and accessible.

In the 2014 survey of institutional research publication repositories (RPRs), more than half the surveyed institutions provided a data storage service, but most library RPRs were not integrated with intuitional data management systems. Results did show that RPRs are increasingly storing meta-data, a trend encouraged by the existence of Australian National Data Service (ANDS) and its leadership on all areas of data management (CAUL, 2014).

If Australia is to continue providing world leading data management infrastructure, greater capabilities are required. Researchers must be equipped with the capability to 'publish' their data – to enable access, management and standardisation across Australia and internationally. For example, Research Data Australia (RDA) enables data to be findable but does not make it accessible, interoperable or reusable through standardisation. The research data should also be linked to publications, to ensure that the reproducibility of results can be achieved in the best possible way. Libraries can provide advice and infrastructure to build links between data and publications, but this requires investment in systems and services for this to work appropriately on a national scale.

This is an opportunity for Australia to provide world-leading infrastructure, and provide substantial efficiencies to Australian researchers. Commercial models like Figshare[6] could provide the basis for a national approach. An approach needs to cover the whole research lifecycle from the planning to the creation and publishing of research data. Significant gaps exist in the provision of infrastructure (not just storage) to help researchers organise, describe and manage data during a project, providing a "curate-as-you-go" environment. This environment should capture provenance, licensing, methodologies, research design, environmental factors and other elements that are often

---

[6] The figShare service: https://figshare.com/

missing when a body of data is ready for publishing or provided as a set of data for re-use by others. Without a national approach Australia risks creating an environment that encourages local and individual solutions that will likely be inconsistent and incompatible.

## (3) Digitisation, imaging and visualisation

CAUL supports the need for national digitisation, imaging and visualisation infrastructure identified in the issues paper. National digitisation capabilities, along with imaging and visualisation technology, are essential capabilities but when funded through various sources offer limited cohesion of existing approaches.

In particular, there has been limited systematic digitisation of important collections and infrastructure to support large-scale ongoing digitisation programs is not yet in place. University libraries and many other institutions are in the process of digitising their collections, but enabling digitisation at scale particularly requires the support of national infrastructure. A national, cohesive approach is required.

In response to the 2011 Strategic Roadmap for Australian Research Infrastructure Exposure Draft[7] CAUL has previously recommended that a decision needs to be made about which model or models are appropriate for a national approach that builds this capability in Australia. Three potential models are potentially worth considering, a distributed/institutional or networked model, a collection-centric model, or a "centres of excellence" model.

An important consideration is that the process of digitising physical objects requires physical access to the object in its primary location. A networked approach with geographically distributed nodes of digitisation equipment and required skills could accommodate these requirements. Post-capture access and preservation of digitised objects could be enabled through the distributed network, or enabled centrally and made available in a distributed manner.

CAUL maintains that any investment of resources must concentrate on the establishment of equipment, the resolution of technical issues and the development of skills to support national digitisation capabilities. Given the overlap of the outputs of digitisation with data management – storage and collections – it is essential that any approach be appropriately connected to ANDS, Research Data Storage Infrastructure (RDSI)[8] and other relevant infrastructure.

## (4) Skills and literacies

Ensuring that researchers have access to the skills required to conduct excellent research is a critical capability issue for Australia. The pervasiveness of new digital technologies in research practice means that new skills are essential for researchers to operate effectively in this environment. For example in HASS disciplines the increasing engagement of researchers with digital infrastructure

---

[7] CAUL response to the 2011 Strategic Roadmap for Australian Research Infrastructure Exposure Draft. Retrieved from:
http://www.innovation.gov.au/SCIENCE/RESEARCHINFRASTRUCTURE/Pages/default.aspx

[8] Research Data Storage Infrastructure: https://www.rds.edu.au

means that researchers require the capability to access and use digital infrastructure to produce excellent research.

However, research workforce skills are not an infrastructure issue as much as a capabilities issue. The idea of an 'Infrastructure ready' researcher or technical specialist proposed by the issues paper implies that the responsibility to be 'ready' lies with the individual researcher. Ideally, there should be services in place to ensure that researchers acquire the necessary skills and competencies that enable them to access and effectively use the necessary infrastructure.

To illustrate the point university libraries and information professionals are well recognised for their expertise in the development of information, digital and research literacy skills and already play a vital role in imparting that knowledge and skill to the research community. The competencies that a library provides are wide ranging and include (but are not limited to):

- ❖ Fluency and skill in the use and application of new and emergent technologies;
- ❖ Capability to participate in digital networks;
- ❖ Capability to provide advice and education on the creation, production and attribution of scholarly communications in a range of media;
- ❖ Capability to manage high quality data and to meet compliance requirements; and;
- ❖ Expertise in intellectual property and other legal and moral rights related to copying, storage, modification of content, and the use of specific digital assets.

University libraries in this frame function as research infrastructure, in the same way that universities function as infrastructure to train the next generation of researchers. A networked approach that builds on the existing infrastructure would be more effective in this case than a centralised solution.

# Governance of national research infrastructure

This part of the submission addresses the following questions from the issues paper:

- ***Question 2:*** *Are these governance characteristics appropriate and are there other factors that should be considered for optimal governance for national research infrastructure.*
- ***Question 8:*** *What principles should be applied for access to national research infrastructure, and are there situations when these should not apply?*

CAUL's position is that the focus of research infrastructure governance may usefully be expanded to include what happens to research outputs, particularly to enable the capability [of researchers] to harness the value of research to advance knowledge, solve complex real world problems and stimulate innovation.

To optimise the value of research, all its outputs (data, software, methods, tools and publications) must be findable, accessible, interoperable and reproducible (FAIR) in a way that easily facilitates use, re-use and application. Discoverability and access to all research outputs not only creates the opportunity for validation of research through reproducibility but also provides the foundation for new areas of research, discovery and innovation. If we are to realise the aspirations of better access to data in all its forms, it will be essential that we develop a national ecosystem that links research data, tools and publications.

The CAUL community is acutely aware of both the opportunities and barriers to increasing research opportunities and impact through access to public, private and research data; and to increasing research discoverability, use and impact. Through the work of the Australian National Data Service, in partnership with the CAUL community, Australia has positioned itself as an international leader in research data policy, standards, services and infrastructure but more must be done.

A major focus on data to date has been on data produced by research institutions (e.g. genomics, urban research). However, many of the complex problems (e.g. combating terrorism, climate change adaption, health services reform) require integrating not only diverse data but also obtaining data from other sources such as government agencies, smart devices owned by councils, banks, telcos, insurance companies, etc. There needs to be a national approach to work with the key industry associations to facilitate access to data for research.

Australia's current national research infrastructure governance model has not delivered coherent approaches to long term develop of nationally significant research data standards, processes, tools and assets.

Research data collections, in many cases, lie in trusted sources that are not managed by research institutions – federal, state and local government departments, state archives and the National Library of Australia. Because of a lack of an over-arching national approach to realising the value of our data assets, each agency has its own standards, methods and governance, all of which have an impact on how researchers have access to the data. Nor are researchers easily able to add to the data holdings of these institutions to make the resources potentially richer for the data owning institutions.

CAUL believes that greater national benefit could be derived through strengthened collaboration in data policy development, services, infrastructure and skills development across the research and government data domains.

CAUL recommends adopting a national approach which puts in place policies, standards and practises to:

1. make publicly funded research outputs findable, accessible, interoperable and reusable. Specifically:
    a. make research publications immediately free to read (through a range of different strategies)  at the point of publication;
    b. make research output data as open as possible, as closed as necessary, in accord with  the *Australian Government Public Data Policy Statement*[9]*;*
    c. apply appropriate AusGOAL licences and international metadata standards to research outputs to ensure accessibility, interoperability and reusability;
    d. retain all necessary rights to enable the authorisation of publication and re-use in any format at any time;
    e. encourage and support researchers to make research outputs available in accord with this policy;
2. value and practically support a range of dissemination models, recognising discipline diversity, whilst maintaining a commitment to the principles of this approach;
3. ensure that the application of this policy is in accordance with the *Australian Code for the Responsible Conduct of Research*[10] and other codes of practice which lay out requirements to disseminate research responsibly, ethically and for the benefit of the Australian and international community.

## References

Council of Australian University Librarians. (2014). *2014 Research Publications Repository Survey: Report*. Canberra, Australia: CAUL.

Department of Education and Training (2016). *National Research Infrastructure Capability Issues Paper*. Canberra, Australia: Australian Government. Retrieved from <https://docs.education.gov.au/node/41051>

Johnson, L., Adams Becker, S., Estrada, V. & Freeman, A. (2015). *NMC Horizon Report: 2015 Library Edition*. Austin, Texas: The New Media Consortium. Retrieved from <http://www.nmc.org/publication/nmc-horizon-report-2015-library-edition>

[9] Australian Government Public Data Policy Statement: https://www.dpmc.gov.au/public-data/public-data-policy

[10] Australian Code for the Responsible Conduct of Research: https://www.nhmrc.gov.au/guidelines-publications/r39