

PRACTICE PAPER

The Australian Research Data Commons

Michelle Barker, Ross Wilkinson and Andrew Treloar

Australian Research Data Commons, AU

Corresponding author: Michelle Barker (michelle.barker@ardc.edu.au)

A research data commons can provide researchers with the data and resources necessary to conduct world class research. More than this, a research data commons can be transformational in facilitating change in the way research is conducted, in terms of both research culture and the availability of research data and analytical tools. This paper describes frameworks needed to build a transformational data commons, through examination of the development of the Australian Research Data Commons (ARDC). ARDC was formed in 2018 as part of a 20-year vision to transform Australia's research culture by enabling access to the digital data and eResearch platforms that can significantly enhance research capacity. ARDC is located within both national and international eResearch ecosystems, and its unique positioning must be understood, alongside the achievements of its three predecessor organisations, to understand the niche from which ARDC aims to provide maximum value and impact. Consideration is given to the challenges inherent in both the current Australian ecosystem and beyond, to articulate ARDC's focus going forward. The paper concludes with consideration of the international dimension, drawing on discussions around the development of a global data commons.

Keywords: Australia; eInfrastructure; data; eResearch; commons; collaboration

Introduction

The Australian Research Data Commons (ARDC) is part of a 20-year vision to transform Australia's research culture by enabling access to the digital data and eResearch platforms that can significantly enhance research capacity and impact. A research data commons is a kind of digital commons that brings together data and related resources (storage, compute, models) to enable researchers to conduct and collaborate on world class data-intensive research.

This concept arises from the increasing interest in the last decade in the economics of commons in general, and in the ability of digital goods and services to be part of a *digital* commons. This might be defined in this way:

information and knowledge resources that are collectively created and owned or shared between or among a community and that tend to be non-exclusive, that is, be (generally freely) available to third parties. Thus, they are oriented to favor use and reuse, rather than to exchange as a commodity. Additionally, the community of people building them can intervene in the governing of their interaction processes and of their shared resources (Fuster Morell, 2010, p5).

In addition to enabling researchers to improve research impacts, a research data commons can be *transformational* in facilitating change in the way research is conducted, in terms of both research culture and the availability of research data and analytical tools.

This paper describes frameworks needed to build a transformational data commons, through the lens of ARDC. Beginning with exploration of the history and aims of ARDC in the larger context of other national and international work on data commons, the paper then considers the challenges inherent in both the current Australian ecosystem and beyond, to articulate ARDC's focus going forward. The paper concludes with consideration of the international dimension, drawing on discussions around the development of a global data commons.

Towards an Australian Research Data Commons

ARDC is part of a 20-year agenda to transform Australia's research culture by enabling access to the digital data and eResearch platforms that can significantly enhance research capacity and outcomes. This can be achieved through changes to the way in which research occurs, and the broader research culture that enables this, and requires engagement with key stakeholders throughout the sector to achieve the goal of creating an Australian research data commons together. This section examines the ecosystem that ARDC functions within, and its aims within this.

The Formation of the ARDC

ARDC is an initiative of the Australian government's National Collaborative Research Infrastructure Strategy (NCRIS), a national network of world-class research infrastructure capabilities that support high-quality research. Capabilities support strategically important research through which Australian researchers and their international partners can address key national and global challenges.

ARDC was formed in July 2018, integrating three existing NCRIS capabilities:

- Australian National Data Service (ANDS)
- National eResearch Collaboration Tools and Resources (Nectar)
- Research Data Services (RDS).

These three organisations contributed variously to significant transformations in the research sector for the last seven to ten years (due to differing starting dates), positioning ARDC well to continue leadership for the second half of this 20-year journey. ANDS was responsible for fundamentally changing how institutions see their responsibilities for the data their researchers produce (ANDS, 2017), including a highly successful data initiative that ensures researchers, research institutions and the nation have access to Findable, Accessible, Interoperable, Re-usable (FAIR) research data. Nectar facilitated investments that enabled researchers to collaborate in using data and tools in new ways, through access to both virtual laboratories and cloud compute (Nectar, 2016). RDS enabled a suite of high-quality data services, including storage and identity services, that emphasised partnerships with eInfrastructure providers to provide reliable solutions (RDS, 2017). Each of these organisations played an important role in community development for research data professionals, domain communities, and research service providers.

These past investments enabled communities, policy, skills, institutional capacity, national data infrastructure coherence, and national services to be addressed together. By ensuring there was corresponding investments in collaboration, storage, computation and domain services, it was possible to have ambitious goals set: all research institutions took part in developing capacity, major funders encouraged data management, and Australia became an active partner internationally in discussions about the role of research infrastructure.

The integration of these organisations was an outcome of National Research Infrastructure Roadmap recommendations to create an Australian Research Data Cloud, which would “deliver a more integrated, coherent and reliable system to meet the needs of data-intensive, cross-disciplinary and global collaborative research” (Department of Education and Training, 2016, p5). The Roadmap notes, “in terms of established infrastructure, the demand for computation and connectivity will continue unabated” and the “unprecedented growth in data volume and complexity places increased demand on infrastructure and the skilled staff needed to support it” (Department of Education and Training, 2016, p27). These recommendations were inspired, in part, by work taking place in developing the European Open Science Cloud (Department of Education and Training, 2016, p28).

The Australian Ecosystem

ARDC is located within both national and international eResearch ecosystems, and its unique positioning must be understood, alongside the achievements of its three predecessor organisations, to understand the niche from which ARDC can provide maximum value and impact.

The value of national infrastructure in digital data and eResearch platforms is well understood in Australia, as evidenced in the work of a number of recent reviews:

- Research Infrastructure Investment Plan (Australian Government, 2018)
- National Research Infrastructure Roadmap (Department of Education and Training, 2016)
- eResearch Framework (Francis, 2016)
- Status Report on the NCRIS eResearch Capability (Cochrane, 2014)

These reviews each concluded that the ability to perform complex computations rapidly, coupled with data storage, complex analytics and data mobility, is essential if Australia is to effectively provide and efficiently take advantage of an evolving data-intensive research environment. The collaborative emphasis of NCRIS capabilities is also seen as crucial. For example, the Status Report on the NCRIS eResearch Capability states: "The national eResearch investments have been timely and critical to the advance and competitiveness of Australian research. ...The deliberate strategy of fostering and implementing collaborative approaches has been a hallmark of the development as acknowledged across a range of community responses" (Cochrane, 2014, p4).

ARDC functions within a broader eResearch infrastructure ecosystem in Australia that includes significant national investments by NCRIS capabilities, national service providers including Australia's Academic and Research Network (AARNet) and Australian Access Federation (AAF), Publicly Funded Research Organisations (PFRs), research institutions and others. These organisations recognise the value of digital data and eResearch platforms and continue to invest heavily in eInfrastructure to support achievement of their strategic research aims. Many of these organisations have their own eInfrastructure strategies, sometimes with budgets that are more substantial than ARDC's. Australia's state-based eResearch service providers also have made a significant investment in infrastructure and services.

The diversity and number of organisations involved in the digital data and eResearch platforms ecosystem in Australia provides both richness and complexity to the landscape. Council of Australasian University Directors of Information Technology (CAUDIT) figures reveal that the sector invests \$2.56 billion annually on IT in higher education and research in Australia and New Zealand (CAUDIT, 2018), demonstrating the need for collaboration across the sector to maximise the benefits of these investments. While some of these organisations focus on institutional, regional and/or research community interests, many have addressed national interests, and have had significant prior collaborations with ANDS, Nectar and RDS. There are a range of other research sector organisations that are also important in this landscape, including federal and state governments, research communities, funders, peak bodies, research academies, a range of other organisations that affect policy (including publishers) and workforce development organisations.

International Partnerships

International concerns about research data have evolved significantly over the last decade (Borgman 2015) and the underlying drivers have changed over time. Initially it was to ensure that data was effectively managed, and then it became to derive the maximum value of data to researchers, research institutions, and the nation. Most recently it has had to support the transformation of research to provide routinely high quality outputs that are reproducible and able to be used to translate the value of research beyond its initial target. This requires attention not just to data storage, but also to curation and connections to software and compute resources.

As a significant but small contributor to international research, Australia has a long-standing tradition of international research collaboration, and has corresponding strengths in international research infrastructure collaborations. This is notable in research areas including Antarctic research, marine research, and astronomy, and has seen major international co-investment in projects such as the Square Kilometre Array. Australia has also contributed to a number of international digital data and eResearch platform initiatives: a major role in establishing the Research Data Alliance; contributions to the international development of virtual laboratories, science gateways and Virtual Research Environments (Barker et al., 2019); development of strong partnerships on research clouds; and building of national research data librarian networks that are international exemplars. This international leadership role continues, with ARDC now driving strong partnerships on broader identification technologies, including supporting the development of Research Activity Identifiers (RAiD) to provide a robust mechanism for identifying where research activities are located, who is collaborating and what resources are being used.

Another example is co-founding the recently formed Research Software Alliance, which promotes research software as a fundamental and vital component of global research. International collaborations also aim influence policy development, to ensure that national investments are optimised. This continues to occur through a number of forums, including OECD Global Science Forum priorities such as the recent Expert Group on Coordination and Support of International Data Networks (OECD 2017) and current Expert Groups on Digital Skills for Data-Intensive Science.

Australia's digital data and eResearch platforms initiatives will continue to benefit from increased alignment with a range of leading international initiatives. There are many significant international initiatives that can provide sustainable solutions and/or best practice to the Australian community, in addition to

those areas that Australian organisations lead in. An increasing number of international programs are also developing to coordinate within and between specific geographic or research discipline communities, such as the European Open Science Cloud (EOSC) and the National Institutes of Health (NIH) Data Commons. Whilst different parts of the Australian research sector engage with these, alignment with national interests and coordination across national initiatives could benefit from more coherent facilitation.

Informing a Research Data Commons

Australia's work over the last 10 years towards the development of a data commons has yielded a number of insights that will influence further work. This includes key areas such as partnerships and identification of their relative responsibilities. These themes have long been identified as relevant, as articulated in this early vision of the Australian data commons:

It is widely understood that the scale of the challenges around digital tools and resources requires that responsibilities be distributed across multiple entities and partnerships, with different roles and responsibilities. It is therefore an appropriate role of a commons to assist institutions and research communities, to join a federation/framework, and to operate together. This often involves developing a framework (or frameworks) that enable silos to engage with a broader community, enabling collective national investment to be made in ways that align with overarching and shared strategic goals. The framework should be developed collaboratively and help to make explicit the roles and responsibilities of the key players in the research ecosystem. (ANDS Technical Working Group, 2007)

The key learnings from the first ten years of Australia's journey revolve around the social and technological changes needed to enable transformation:

- **Policy and incentives:** There must be a close relationship between policy and the development of digital data and platforms to facilitate cultural change. This can range from the way in which the policy of funders influence data sharing behaviours, to how institutional human resources policies affect the career development of research support professionals such as research software engineers. In establishing incentives to drive cultural change it is important to understand that value differs – the value to the researcher is not the same as to the research institution, and both are important.
- **Frameworks and standards:** Data is not intrinsically valuable – additional work is needed to make it as FAIR as possible, to ensure that data is as high quality as possible, and to situate it as a trusted research input and output (Borgman 2015).
- **Community and culture:** Community development is needed to ensure sustainability of solutions and change culture. It is vital that all parties to understand who has responsibility for what, in both the short and long-term. This should include a focus on the role of research institutions and international partnerships.
- **Skilled workforce:** This is crucial for meeting the challenges required to excel in the new digital economy. Effective and sustainable digital infrastructure requires workforce development programs, supported by sustainable communities, to create an environment that supports and capitalises on the data collections, collaborative platforms and underlying infrastructure.
- **Platforms and tools:** Online analytical platforms and tools are needed to enable the delivery of world-leading informatics capabilities for researchers and their collaborators. The provision of collaboration environments is invaluable.
- **Underlying infrastructure:** A strong foundation providing data storage and compute is essential. This infrastructure provides the base on which collaborative platforms can be used to analyse data, to maximise the value of data assets.

The Aims of the Australian Research Data Commons

ARDC is a transformational sector-wide initiative, working with partners to build a coherent, national and collaborative research data commons which delivers a world-leading data advantage, facilitates innovation, fosters collaboration for borderless research, and enhances researchers' ability to translate their work into outputs for a better world.

ARDC seeks to facilitate national coordination of digital data and eResearch platforms initiatives, acting as a catalyst to enable the required social and technical infrastructure. This will occur through discussions and alignment of strategic plans with key stakeholders including the digital data and eResearch platform

capabilities identified in the National Research Infrastructure Roadmap (Academic and Research Network (AARNet), Australian Access Federation, National Computational Initiative (NCI) and Pawsey Supercomputing Centre. As a driver and enabler within the Australian eResearch ecosystem, ARDC seeks to support and engage with the broader community to create an Australian research data commons.

ARDC can also play a role in facilitating alignment with other commons initiatives internationally. The 2018 Research Data Alliance plenaries included sessions to advance collaboration between different international initiatives, including the African Open Science Platform, ARDC, European Open Science Cloud (EOSC), National Institutes of Health (NIH) Data Commons, and Canadian approaches. There is interest in convergence on a framework to facilitate alignment, efficiency and interoperability where possible and desirable. The European Commission has also identified as one of the five priority deliverables for EOSC to ensure EOSC activities align and coordinate with others globally.

Australia faces similar challenges to national and international initiatives focused on developing research data commons, in assisting communities with separate, discipline specific data commons to engage within this broader framework. This engagement is vital for ensuring that collective national investments in research data occur in ways that maximise discoverability and reuse. Integration of existing commons approaches is complex, with achievement of the vision of a research data commons depending on both significant socio-cultural and technological changes. **Figure 1** illustrates ARDC's framing of these, which encompass the lessons learned from ARDC's predecessor organisations.

ARDC's Strategic Plan 2019–23 describes the five strategic themes that frame the implementation of ARDC's vision:

1. Coordination and Coherence: Facilitating an Australian research data commons
2. People and Policy: Transforming culture and community
3. Data and Services: Maximising the value of Australia's data assets
4. Software and Platforms: Enabling research insights & supporting collaboration
5. Storage and Compute: Providing foundation infrastructure (ARDC, 2019)

The strategic plan also details the impact framework methodology utilised by ARDC to tie its activities to societal impact.

In line with Australia's learnings from previous work in this area, ARDC identifies partnerships as key to its mission, with responsibility frameworks needed to articulate roles and responsibilities, to ensure development of sustainable frameworks. While ARDC will function as a coordinating organisation to motivate and energise this activity, the commons itself must be a collaborative endeavour, developed in concert with broader data issues and interests. ARDC's engagement strategy aims to facilitate effective engagement at the appropriate level with strategic stakeholders of all types throughout the sector, to achieve the transformational goals shared by both ARDC and the larger community. Given the way responsibilities are shared



Figure 1: Key focus areas required to bring about an effective research data commons (inspired by a diagram from the Center for Open Science).

across the participants in the research ecology, those programs will also have to have a suitable impact on researchers, research administrators, research support staff, research institutions, research funding agencies and government departments, if the goal is to be achieved.

Towards a Global Data Commons

Digital infrastructure lies at the heart of changing practices in research. ARDC aims to continue to facilitate both national and international increases in research impacts through ongoing development of Australian digital data and eResearch platform capabilities, in alignment with discipline-specific and international initiatives.

ARDC is uniquely positioned in the Australian ecosystem to provide coherency at the national level, by facilitating the development of an Australian research data commons. Extensive engagement with key stakeholders throughout the sector is needed to achieve the goal of creating an Australian research data commons, together, that represents the needs of the sector. Discussion is also needed on both the vision and aims of the common, alongside mechanisms to establish sustainability infrastructure that is also flexible enough to adapt to future needs.

ARDC is also committed to the coordination and collaboration needed to work towards a global data commons, to ensure that the various solutions under development align and interoperate to meet researchers' needs. Open requires more than sharing data and artefacts, it also needs a culture of sharing practice.

The result of work by the ARDC over the next 10 years will be to expand international, institutional, service provider and research partnerships through compatible infrastructure, to enable Australia to build national and international partnerships in support of the Sustainable Development Goals.

Competing Interests

The authors are employees of the Australian Research Data Commons.

References

- ANDS.** 2017. *Data Impact*. https://www.ands.org.au/news-and-events/dataimpact?utm_source=home-page-carousel&utm_medium=Web-banner&utm_campaign=data-impact-ebook-published [Last accessed 23 January 2019].
- ANDS Technical Working Group.** 2007. *Towards an Australian Data Commons: A proposal for the Australian National Data Service*. <https://web.archive.org/web/20091024214417/http://www.pfc.org.au/pub/Main/Data/TowardstheAustralianDataCommons.pdf> [Last accessed 30 January 2019].
- ARDC.** 2019. Strategic Plan 2019–23. <https://ardc.edu.au/wp-content/uploads/2019/05/ARDC-Strategic-Plan-2019-2023.pdf> [Last accessed 30 July 2019].
- Australian Government.** 2018. Australian Government Research Infrastructure Investment Plan. https://docs.education.gov.au/system/files/doc/other/ed18-0069_-_he_-_research_infrastructure_investment_plan_-_public_version.pdf [Last accessed 22 January 2019].
- Barker, M, Delgado Olabarriga, S,** et al. 2019. The global impact of science gateways, virtual research environments and virtual laboratories. *Future Generation Computer Systems*, 240–248. <https://www.sciencedirect.com/science/article/pii/S0167739X18314018?via%3Dihub> [Last accessed 17 January 2019]. DOI: <https://doi.org/10.1016/j.future.2018.12.026>
- Borgman, C.** 2015. *Big Data, Little Data, No Data: Scholarship in the Networked World*. The MIT Press. DOI: <https://doi.org/10.7551/mitpress/9963.001.0001>
- CAUDIT.** 2018. *Benchmarking Report Summary*. https://www.caudit.edu.au/system/files/Media%20library/Resources%20and%20Files/Benchmarking%20and%20Analytics/Infographic_2017_web%20v1.1.20181102.pdf [Last accessed 17 January 2019].
- Cochrane, T.** 2014. *Status Report on the NCRIS eResearch Capability*. https://docs.education.gov.au/system/files/doc/other/abridged_eresearch_status_report_-_web.pdf [Last accessed 19 January 2019].
- Department of Education and Training.** 2016. *National Research Infrastructure Roadmap*. <https://www.education.gov.au/2016-national-research-infrastructure-roadmap> [Last accessed 3 January 2019].
- Francis, R.** 2016. *Towards an eResearch Framework*. <http://aero.edu.au/wp-content/uploads/2015/03/Rhys-eRFramework-AeROForum2016.pdf> [Last accessed 17 January 2019].
- Fuster Morell, M.** 2010. Dissertation: Governance of online creation communities: Provision of infrastructure for the building of digital commons. http://www.onlinecreation.info/?page_id=338 [Last accessed 30 January 2019].

- Nectar.** 2016. *Impact*. https://nectar.org.au/wp-content/uploads/2016/04/NEC_004-Codex_230x260_FINAL-WEB.pdf [Last accessed 23 January 2019].
- OECD.** 2017. "Co-ordination and support of international research data networks". *OECD Science, Technology and Industry Policy Papers, No. 51*. Paris: OECD Publishing. [Last accessed 23 January 2019]. DOI: <https://doi.org/10.1787/e92fa89e-en>
- RDS.** 2017. *Collaborative Projects*. <https://www.rds.edu.au/collaborative-projects> [Last accessed 23 January 2019].

How to cite this article: Barker, M, Wilkinson, R and Treloar, A. 2019. The Australian Research Data Commons. *Data Science Journal*, 18: 44, pp.1–7. DOI: <https://doi.org/10.5334/dsj-2019-044>

Submitted: 18 February 2019

Accepted: 23 August 2019

Published: 05 September 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Data Science Journal* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 