

ARDC's Submission to the ARC ERA EI Public Consultation

19 October 2020

In this submission, the Australian Research Data Commons (ARDC) offers responses to selected questions drawing upon our expertise and experience in the field of research data and e-infrastructure. (For information regarding the work of the ARDC, please see the end of this document.) Our submission examines three issues: (i) an ERA objective and contextual indicators; (ii) research data as an assessable output with potential for further impacts; and (iii) technological applications involving the use of persistent identifiers to minimise the reporting burden for higher education institutions. In addition, we discuss how EI impact narratives have been utilised in a study commissioned by the ARDC to acquire insights into the role of data in the context of research yielding non-academic impacts.

ERA Policy

Value of ERA

3.1 To what extent is ERA meeting its objectives to:

c. Identify excellence across the full spectrum of research performance. *A very large amount; A large amount; A moderate amount; A small amount; Not at all. Please explain your answer.*

The ERA measures the excellence of Australian research primarily by measuring the quality of research outputs with some consideration as to staff and income inputs.

The ERA measures do not adequately account for the quality of the environments in which researchers work as a determinant of the quality of outputs able to be achieved.

As such, the ERA is missing a critical lead indicator for the expected future performance of Australian researchers. Without it, decision makers will have insufficient information to assess and thereby influence the quality of future outputs.

Previously, assessing the environment in which researchers worked may have consisted of measuring access to and quality of physical equipment and facilities.

Increasingly, assessing the environment must account for access to and utilisation of eResearch infrastructure both within and between institutions. As indicated in the 2015 Clark Review¹, 'excellent research requires excellent research infrastructure'.

In 2016, Lord Nicholas Stern conducted an independent review of university funding in the United Kingdom (UK). Lord Stern observed the need for an improved assessment of the environment in which research was conducted, stating:

¹ 'Research Infrastructure Review 2015 (Clark Review)', (p. 21).

Research is a long-term process that requires commitment and support. The purpose of the environment assessment is to encourage and reward institutions which endeavour to develop the vitality and sustainability of their research environment, including its contribution to the wider discipline or research base².

As a result of this finding, the UK is piloting an environmental assessment in 2021, as part of their Research Excellence Framework³.

The ARDC recommends the ARC consider an environmental assessment as part of the ERA, specifically including an assessment of eResearch infrastructure.

ERA Methodology

Contextual Indicators

3.19 The volume and activity indicators are still relevant to ERA. *Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree*. Please explain your answer.

The ERA Submission Guidelines outline the ‘eligible research output types’, which include both traditional (para 4.4.8) and non-traditional (para 4.4.9) types of research output.

Of note, neither of these categories includes ‘digital artefacts’ (such as software and research data) as eligible research output types. Where it could conceivably cover some ‘digital artefacts’⁴, many ForCodes are excluded specifically from including this as an accessible output type⁵.

This is in contrast to the UK Research Evaluation Framework (REF)⁶, which accepts a range of digital artefacts including:

- Software
- Website content
- Digital or visual media, and
- Research data sets and databases.

Research data and research software across all FOR Codes is regarded in Australia and internationally as a legitimate and highly valued research output.⁷ Indeed in some disciplines depositing data outputs in domain repositories is considered as such an investment in future

² Stern, N (2016) ‘Research Excellence Framework Review: An independent review of university research funding by Lord Nicholas Stern’, (p, 16). Accessed 8 October 2020.

<https://www.gov.uk/government/publications/research-excellence-framework-review>.

³ <<https://www.ref.ac.uk/news/ref-environment-pilot-guidance-and-criteria-published/>>.

⁴ Such as for ‘4.4.9.4 Recorded/rendered creative works’ of the ERA 2018 Submission Guidelines, p.37.

⁵ Discipline Matrix ERA 2018.

⁶ ‘REF2020: Guidance-on-Submissions’, pp. 110-112. Accessed 8 October 2020.

https://www.ref.ac.uk/media/1447/ref-2019_01-guidance-on-submissions.pdf.

⁷ For example, Australian Government Productivity Commission “Data Availability and Use” (2017) pp 115-118.

research that the research leadership has made it a prerequisite for journal publication and grant awards.⁸

In recognition of this, the Group of Eight signed the ‘Sorbonne Declaration on Research Data Rights’⁹ in January 2020, which committed them to ‘ensuring that peer-reviewed publications are supported by the necessary FAIR data sets as research findings must be accessible, verifiable and replicable’.

This commitment, amongst others in the declaration, was considered by the signatories as ‘an essential issue for the quality and transparency of research’ as well as necessary ‘to accelerate scientific discoveries and economic development’.

The ARDC recommends the ARC expand the eligible research output types to include digital artefacts in a manner comparable to the Research Evaluation Framework of the UK. Specifically, this should include both software and data as legitimate outputs of research activity.

One might legitimately ask, “Would the inclusion of new information types include further reporting burden?” Not if we take advantage of contemporary information systems.

Information about data outputs are already being collected in a number of systems, for example:

- ORCID records already contain data outputs (it is one of the registered types of “work”).
- CrossRef collects references to any datasets cited in article references in a comprehensive global service called EventData.
- The ARC collects “data outputs” in the final reports of ARC-funded projects.
- The ARDC has a national catalog¹⁰ of research data outputs with currently some 150,00 collections registered (these descriptions are in turn mostly harvested automatically from institutional data services).

All the above information sources (and many more) are queryable through APIs to pre-populate a data output report.

Granted, none of those sources would provide a pre-existing comprehensive ERA/EI-ready information source for research data outputs; nevertheless they do show that the broader ecosystem is already collecting this information. Researchers and research organisations are already providing this information and these information pathways would simply scale up to meet the needs of ERA/EI.

⁸ For example, “publication of new macromolecular structures in most scientific journals is contingent on mandatory deposition to the PDB of the 3D atomic coordinates comprising the structural model plus experimental data used to derive the structures and associated metadata. Many governmental and non-governmental research funders also require PDB deposition of unpublished macromolecular structure data”. <<https://doi.org/10.1093/nar/gky949>>

⁹ <<https://www.leru.org/files/Sorbonne-declaration.pdf>>.

¹⁰ Research Data Australia <<https://researchdata.edu.au/>>.

In any case, it might be wise to make new output types like data and databases optional so that fields of research that are data-intensive can avail themselves of the option, and those that are not don't have to.

Further details on the emerging global information systems that would support the gathering of this information are given in Q5.10 below.

The ARDC continues to provide support for all Australian universities to “publish data”. Together the ARC and ARDC could create a comprehensive and quality overview of Australia’s valuable research data outputs. Such a comprehensive inventory would be a powerful policy side effect of the ERA/EI process, underpinning the systematic integrity and reproducibility of research as well as innovative re-use.

Below, the ARDC refers to a study that has carefully analysed the 2018 EI impact narratives and established the unique pathway to impact that research data represents. A coherent ERA/EI framework would include quality research data as a research output due to its demonstrated ability to support both further excellent research¹¹ and broader impacts¹².

EI Overview

Q4.2 The EI objectives are appropriate for the future needs of its stakeholders. *Strongly agree; Agree; Neither agree or disagree; Disagree; Strongly disagree. Please explain your answer.*

Agree.

In its submission for the EI 2018 Public Consultation, the Australian National Data Service (ANDS) - one of three organisations preceding ARDC - remarked that it supported the proposed EI objectives as set out in the ARC’s Draft Submission Guidelines. The ARDC affirms this position. In making our submission for this consultation, we acknowledge that metrics and indicators do not operate in neutral ways. As such, we appreciate the value of the ERA EI Review and the opportunity to contribute. Clearly, stakeholder feedback and survey responses will offer valuable insights on areas of public consensus and contestation. Public input will also assist the ARC to improve its assessment processes, review the implementation of research evaluation objectives, and effectively respond to any negative or unintended consequences upon the conduct of research in the Australian higher education sector.

Q4.3 What impact has EI had on:

- the Australian university sector as a whole? *Please describe.*
- Individual universities. *Please describe.*
- researchers. *Please describe.*
- other sectors outside of academia? *Please describe.*

¹¹ See for example: “The High Impact of Astronomical Data Archives”
 <<http://www8.nationalacademies.org/astro2010/DetailFileDisplay.aspx?id=423>>.

¹² Eric A Jensen and Mark Reed, *Investigating the Link Between Research Data and Impact (Phase II)* (ARDC Commissioned Report, 2020) <<https://doi.org/10.5281/zenodo.3543505>>.

For the ARDC, the 2018 EI Assessment has led to a project involving the analysis of EI outcomes and conversations with our stakeholders regarding the results from our project (details below). The corpus of impact narratives has formed a valuable dataset in its own right. The ARDC impact analysis project would not have been possible without the EI corpus. The EI process has made research impact a mainstream conversation at research institutions and has set the scene for conversations with the sector on how infrastructure contributes to impact.

Q4.4 How do you, or your organisation, use EI outcomes? *Please describe.*

Q4.5 The EI outcomes are valuable to you or your organisation. *Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.*

Q4.6 How else could EI outcomes be used? *Please describe.*

Response to Q4.5: Agree.

Response to Q4.4, Q4.5 and Q4.6

The ARDC has commissioned a study to investigate the impact narratives from the 2018 EI Assessment to explore the relationship between research data and non-academic impacts. This study was led by Associate Professor Eric Jensen from Warwick University. Our study centres on the following question:

In the most explicitly data-focused EI impact cases, how does research data contribute to the production of social, economic, environmental, and cultural benefits (“research impact”) to non-academic end-users of research within the impact narratives?

Related to work on the 2018 EI outcomes, the commissioned study also reviewed the outcomes from the UK 2014 Research Excellence Framework. Results from the analysis of UK and Australian case studies are ‘broadly similar’ indicating ‘structurally parallel patterns’ in terms of how research data is used for non-academic impacts in the international context.¹³ Our project involved the participation of colleagues from the ARC, NHMRC, CSIRO and the Research Data Archive. The results from this study will inform our work on how the ARDC could grow the Australian Research Data Commons as a resource for the benefit of researchers and the wider community. Our work aligns with federal government policy developed over a decade ago to establish the ‘Quality and Accessibility Frameworks for Publicly Funded Research’ as part of a revised initiative entitled, ‘Backing Australia’s Ability - Building our Future through Science and Innovation.’¹⁴ The work of making visible the role of data outputs in research is in line with the spirit of the 2004 Accessibility Framework.

¹³ Eric A Jensen and Mark Reed, *Investigating the Link Between Research Data and Impact (Phase II)* (ARDC Commissioned Report, 2020) 4.

¹⁴ Department of Education, Science and Training, *Quality and Accessibility Frameworks for Publicly Funded Research, Backing Australia’s Ability – Building Our Future through Science and Innovation* (2004) <https://web.archive.org/web/20061004093612/http://backingaus.innovation.gov.au/2004/research/qual_pub_res.htm>.

From our perspective, the utility of EI outcomes could be enhanced if there was a systematic method to collect and report on the use of research data and datasets documented in submissions for EI Assessments. Our interest in data-focused EI outcomes is supported by research results from our commissioned study. The study found that research data plays an ‘essential role’ in delivering a range of impacts of value to society including contributions to the economy, public health, crime prevention, the environment, and government administration and regulation. The 2018 EI outcomes provided empirical evidence affirming that research data is a pathway to engagement with, and benefits for, non-academic end-users of research. Importantly, the study pointed to the unique role of research data. Jensen and Reed observed that 93% of identified impacts in the EI 2018 case studies would not have been possible without supporting research data. Alternatively put, research data provides a ‘unique pathway to achieving an identified impact’ in a large number of case studies.¹⁵ Jensen and Reed advises that through capacity building initiatives involving government, industry and universities, additional impacts could be acquired by re-using or adapting existing research data for ‘new purposes.’¹⁶ This work would require the provision of access to research data and support in the interpretation and analysis of such data. The findings from our commissioned study share resonances with a common theme we deduced from the companion case studies to the Watt Review.¹⁷ From a number of these case studies, it can be shown that the sharing of research data played a key enabling role to facilitate cross-sector engagement for university-industry collaborations. Similarly, the Department of Education, Skills and Employment has prepared case studies from the National Collaborative Research Infrastructure Strategy (NCRIS) illustrating the critical contributions of research data in university-industry partnerships across several fields including grain production, steel manufacturing and indigenous health.¹⁸

Given the potential value which can be leveraged from existing research data collections, we propose that the production of research data and datasets from publicly funded research should be regarded as research outputs or evidence of research related activities for the purposes of EI assessment. Due recognition provided for the collection and curation of research data could serve as an incentive for researchers to share their data. Specifically, in future EI Submission Guidelines, we recommend that the ARC list data or datasets in the examples of engagement and pathway to impact. We note the list of examples in the 2018 EI Submissions Guidelines - ‘Appendix F - Examples of additional quantitative information for the engagement narrative’. Our proposed examples for this list include ‘the number of datasets made available to government or industry partners’ and ‘the number of data accesses by government or industry partners’. While Appendix F is designed as guidance for the preparation of engagement narratives, there does not appear to be an equivalent list for impact narratives. As such, we propose that the ARC address this omission with a comparable list of example impact pathways. Such a list should also include research data and datasets. In addition, we note that the ARC had circulated a copy of the Draft 2018 EI Submission

¹⁵ Eric A Jensen and Mark Reed, *Investigating the Link Between Research Data and Impact (Phase II)* (ARDC Commissioned Report, 2020) 25.

¹⁶ As above, 4.

¹⁷ Ian Watt, *Case studies on University-Business Collaborations: Review of Research Policy and Funding Arrangements* (November 2015), Department of Education, Skills and Employment <<https://docs.education.gov.au/node/38981>>.

¹⁸ Department of Education, Skills and Employment, *NCRIS Case Studies*, <<https://www.education.gov.au/ncris-case-studies>>.

Guidelines for public comment. We support this practice. If the ARC were to conduct a similar consultation on guidelines for the next EI Assessment, we would be interested in providing input and offering suggestions of specific examples involving the use of research data.

Recommendation:

That the ARC develops a list of examples of research impacts and impact pathways similar to the one set out in Appendix F for engagement narratives. Both lists should incorporate examples on research data and datasets due to the well evidenced position that research data is a pathway to engagement and impact.

Utilising technological advances and pre-existing data sources

Q5.7 ORCID iDs should be mandatory for ERA.

Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree.

Please explain your answer.

a. What are the advantages and/or disadvantages? *Please explain your answer.*

Q5.8 The automatic harvesting of output data using ORCID iDs would streamline a university's submission process.

Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.

a. What are the advantages and/or disadvantages? *Please explain your answer*

Response to Q5.7: Strongly agree.

Response to Q5.8: Strongly agree.

Response to Q5.7a and 5.8a

ORCID identifiers have become essential tools for researchers across all disciplines. By connecting researchers with their publications, affiliations, grants, data and other research outputs, ORCIDs can enhance researcher visibility, simplify the publication process and improve the streamlining of workflows through integrated auto-population tools. Since 2014 the ARDC, through its predecessor ANDS, has worked with the ARC, NHMRC, Australasian Research Management Society (ARMS), Council of Australian University Librarians (CAUL), Council of Australasian University Directors of Information Technology (CAUDIT) and other groups in the Australian research sector to support the adoption of ORCID as a standard identifier for Australian researchers. ARDC is a founding member of the Australian ORCID Consortium of which almost all Australian Universities are members. We have also contributed to the drafting of the ORCID 2020 Vision Statement that specifies all Australian researchers have an ORCID iD and use of that identifier in ARC and NHMRC grant systems. The sector-wide approach to ORCID adoption through the Consortium has put us in a position to be able to mandate ORCIDs for ERA submission. Including ORCIDs as a mandatory part of the submission process is consistent with the decision of the sector to adopt ORCID and will assist to streamline the submission process for both researchers and the ARC, saving time for both parties and supporting attribution and credit.

We speculate that it may be possible to mandate the use of ORCID IDs since the next round of ERA will not take place until 2023. Obviously, if the ARC were to introduce such a policy, it will be desirable to provide reasonable notice in order for researchers to undertake the required tasks for compliance. However, the groundwork has already been laid through the work of Australian research institutions via the ORCID Consortium. Given its work, the Australian ORCID Consortium would be in an ideal position to offer valuable input and advice on policy implementation if the ARC were to mandate the use of ORCID identifiers for ERA and EI assessment.

Q5.9 DOIs should be mandatory for ERA. *Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree. Please explain your answer.*

a. What are the advantages or disadvantages? *Please explain your answer.*

Agree.

DOIs have become an important component of publication metadata. Similarly to ORCID iDs, the use of DOIs could improve access to research outputs, enhance content management, and promote data sharing within and beyond the scholarly ecosystem. One key benefit of DOIs is access to citation data and other types of tracking metrics. Such data can offer information on how a particular piece of research output is used or referenced across the internet.

ORCID technologies permit researchers to link or import DOIs into their ORCID record. Since the ARC's Research Management System (RMS) can be connected to ORCID IDs, it follows that researchers can also submit DOIs to the RMS.

In question 3.19 above the ARDC recommends the addition of additional research output types following the UK example. Were the ARC to do that, then DOI will also be applicable to a number of "digital object" outputs such as datasets, software, algorithms, videos, etc. The DOI system caters for extensible object types, and major DOI registration agencies, such as CrossRef and DataCite, have extended the DOI system to all these kinds of digital objects.

Notwithstanding the pervasive use of DOIs, it is also necessary to acknowledge that DOIs cannot be allocated for all types of research outputs. Some kinds of monographs, government reports, and very dynamic databases for example do not lend themselves to DOI for cultural and technical reasons. Nevertheless, a "default" policy requirement for DOIs could also take into account disciplinary norms and categories of research outputs that cannot be tagged with DOIs.

Q5.10 Are there new ways to collect data to reduce the cost and burden to universities of participating in ERA and EI whilst maintaining the robustness of the ERA and EI process? *Yes/No. Please explain your answer.*

a. What are the advantages and/or disadvantages? *Please explain your answer.*

Yes.

Related to questions 5.8 and 5.9, it should be noted that the use of ORCIDs for researchers and DOI for research outputs (both publication and data) allows EI and ERA to leverage global information systems. ORCID and DOI have demonstrated their ability to reduce data input time in the ARC's grant application process and the same is true for ERA.

The interconnectedness of information systems continues to grow and will continue to make ERA/EI reporting easier. As an example, the PID Graph has been developed by a European Commission funded project of which the ARDC is one of the partners. This initiative offers enormous potential in tracking research impact. The graph uses persistent identifiers including ORCIDs and DOIs to link and trace connections between researchers, their research outputs, institutional affiliations and the grants they receive. This is a solid foundation for tracking research impact as it enables rich querying of the data – for example, to find all of the researchers connected with a grant, the data they produced, the publications that ensued and so on. Encouraging sector-wide adoption of the types of identifiers used in the graph such as ORCIDs and DOIs, will enable better data for the PID Graph and, therefore, enhance its ability to be used by Australian research funders and institutions in tracking research impact.

Please also see our response to 3.19 with some innovative ways to collect proposed new information types.

Australian Research Data Commons: 'About Us' Statement

The Australian Research Data Commons (ARDC) is a transformational initiative that enables Australian research community and industry access to nationally significant, leading edge data intensive e-infrastructure, platforms, skills and collections of high-quality data.

In partnerships with organisations, the ARDC leads facilitations that work towards a coherent research environment to enable researchers to find, access, contribute to and effectively use services to maximise research quality and impact.

For more information, please visit our website: <www.ardc.edu.au>.

The contact for this submission is Dr Adrian Burton, Director, Data, Policy and Services
adrian.burton@ardc.edu.au or (02) 6125 6659