Submission: Productivity Inquiry - Australia's Data & Digital Dividend

Rosie Hicks, CEO ARDC

21/10/2022

The Australian Research Data Commons (ARDC) thanks the Productivity Commission for the opportunity to comment on the interim reports of the the *5 Year Productivity Inquiry*.[1]

About ARDC

The ARDC drives the development of national digital research infrastructure that provides Australian researchers with a competitive advantage through data. The ARDC is Australia's peak body for research data. We aim to accelerate research and innovation by driving excellence in the creation, analysis and retention of high-quality data assets. We facilitate access to national digital research infrastructure, platforms, skills, data sets and tools from academia, industry and government for all Australian researchers. The ARDC is funded through the Australian Government's National Collaborative Research Infrastructure Strategy (NCRIS) to support national digital research infrastructure for Australian researchers.

Background

NCRIS facilities were born from various inquiries and strategies acknowledging that 'world-class research infrastructure boosts the productivity of Australia's researchers'[2] and that knowledge generated from research has positive spill-over effects. For example, a 2021 study by Lateral Economics calculated the direct benefit of investment in NCRIS facilities was above a $7 return for every $1 invested.[3]

The *2021 National Research Infrastructure Roadmap*[4] noted that the use and reuse of data is critical. National research infrastructure 'must make data accessible for researchers, and data generated from research within academia also needs to be accessible to government and other users'. As digital research infrastructure is fundamental to Australia's research effort it concluded there was a need for a *National Digital Research Infrastructure Strategy* aimed at coordinating and integrating the national digital research infrastructure ecosystem and underpin collaboration at scale.[5] It was stated that this Strategy

[1] Public inquiry - Productivity Commission
[2] 2011 Strategic Roadmap for Australian Research Infrastructure, p.4.
[3] National Collaborative Research Infrastructure (NCRIS) spending and economic growth by Lateral Economics, 2021
[4] 2021 National Research Infrastructure Roadmap - Department of Education, Australian Government
[5] ibid., Recommendation 7.

should be consistent with, and supportive of, other whole of government initiatives, such as the *Digital Economy Strategy 2030*[6] and the *Australian Data Strategy*.[7] It was also recommended the Strategy should be developed by the government over the next year following the Roadmap with immediate insights feeding into the planned *2022 Research Infrastructure Investment Plan*.

To maximise investments in this area, it is critical that all parts of the national ecosystem support these intended outcomes. It is in this context that the ARDC suggests actions the government can take to support ongoing improvements in research productivity.

Responses to the Interim Reports of the Productivity Inquiry

Given the role of the ARDC, the focus of this submission is primarily on Interim Report Two - *Australia's data and digital dividend*,[8] but it references related issues in other Interim Reports where relevant.

Improved Access to Data Generated as a Result of Public Funding

The productivity gain from 'improved access to data generated as a result of public funding' is highlighted across two of the Interim Reports, specifically in the following sections:

- More value from data provided to government agencies[9]
- More value from data held by government-funded service providers[10]
- Absorptive capacity and the Research and Development (R&D) Tax Incentive[11]
- Access to publicly funded research[12]

That data predominantly funded by taxpayers should be available for use in generating value for the community is well established. As an adherent to numerous recommendations of the OECD Council, Australia has repeatedly agreed over several decades to implement this principle, such as for the:

- Recommendation on Enhancing Access to and Sharing of Data[13]
- Recommendation concerning Access to Research Data from Public Funding[14]
- Recommendation on Environmental Information[15]
- Recommendation on Open Government[16]

---

[6] Digital Economy Strategy
[7] Australian Data Strategy
[8] Interim Report 2 - 5 Year Productivity Inquiry: Australia's data and digital dividend
[9] Interim Report 2, p.47.
[10] Ibid., p.51.
[11] Interim Report 3, p.39. See also para 1.3, p.19.
[12] Ibid., p.78.
[13] Recommendation of the Council on Enhancing Access to and Sharing of Data
[14] Recommendation of the Council concerning Access to Research Data from Public Funding
[15] Recommendation of the Council on Environmental Information
[16] Recommendation of the Council on Open Government

- Recommendation for Enhanced Access and More Effective Use of Public Sector Information[17]

With regard to these and similar commitments, Australia has to make better progress in implementing them across jurisdictions, policy areas and programs nationally if productivity is to improve. In terms of the specific headings in the interim reports, the ARDC makes the following observations:

More value from data provided to government agencies. The Australian Government has recently passed the *Data Availability and Transparency Act 2022* (DAT Act)[18] and established the Office of the National Data Commissioner (ONDC) charged with implementing the DAT Scheme and other measures. This should lead to improvements in the accessibility of data held by government, however:

- As identified in the Productivity Commission's *Inquiry into Data Availability and Use*[19] (and as quoted on the ONDC website[20]) 'You can't use data if you don't know it exists'. Under the DAT Act there is no obligation for government bodies to make public the metadata of data they hold. While the ONDC is currently encouraging agencies to populate an inventory of datasets, agencies are not obligated or adequately resourced to broadly advertise (and then arrange the sharing of) the data they hold. This will inevitably limit discoverability and use, as has occurred in the past. Alternatively, it may result in researchers using much less efficient methods to find and access public sector data, such as provisions in the *Freedom of Information Act 1982* (FOI Act).[21] Ideally, agencies should be obligated (and adequately resourced) to publish publicly the metadata of data they hold. This obligation could be under the DAT Act or else as a specific provision made under the existing Information Publication Scheme of the FOI Act.[22]
- Participation in the DAT Scheme remains voluntary for government agencies as well as across jurisdictions with no intention to progressively transition data sharing arrangements nationally to this Scheme. This risks the Scheme being under-utilised and multiple data sharing frameworks running concurrently, increasing the costs overall for organisations trying to produce value through use of multiple public sector data sets. In contrast there are more robust and pan-jurisdictional data sharing frameworks elsewhere, notably in the European Union (EU).[23]
- Lastly, and as noted in Interim Report 2, the DAT Act 'does not allow for government data to be shared with the private sector, including businesses and not-for-profit organisations'.[24] This is a significant exclusion from the Act, both for researchers in non-university sectors but also for academic researchers collaborating with commercial and not-for-profit organisations in consortia - a characteristic of many national R&D grant programs. While *Australia's Digital Economy*

[17] Recommendation of the Council for Enhanced Access and More Effective Use of Public Sector Information
[18] https://www.legislation.gov.au/Details/C2022A00011
[19] Inquiry report - Data Availability and Use - Productivity Commission, p.159.
[20] Data Discovery | Office of the National Data Commissioner
[21] https://www.legislation.gov.au/Details/C2022C00293
[22] Information Publication Scheme - Home
[23] European Data Governance Act | Shaping Europe's digital future
[24] Interim Report 2, p.50.

Strategy,[25] the *Australian Data Strategy*[26] and the DAT Act all state an intention to share more public sector data across sectors, there remain key gaps and impediments in current schemes that overall constrain the productivity of Australian researchers and therefore users of research.

A 2017 OECD survey established that 74% of scientists reported a 'High' or 'Very High' dependence on public sector information for research.[27] A recent analysis by the Institute for Methods Innovation confirmed that public sector data shared effectively with the research sector creates considerable societal value including to the government itself.[28] However, many public sector data initiatives are designed with either government to government sharing in mind or simplified data products for general public use. Whilst public servants individually are keen for public sector information to empower research, often the structural settings of programs are not supportive. Sometimes access by research is provided only many years later. Our recommendation is that research be a default purpose for public sector data programs and that access by the research sector be a priority.

More value from data held by government-funded service providers. The ARDC supports the recommendation that more value should be gained from data held by government-funded service providers. This should cover the broadest range of providers and contracts. For example, there are co-regulatory arrangements for product labelling[29] and product stewardship schemes[30] managed by the Department of Climate Change, Energy, Environment and Water (DCCEEW). However, there is minimal data published on data.gov.au by either DCCEEW or the co-regulatory bodies. As such, there is minimal data publicly available on the entities being regulated, product drop-off points or material volumes processed. The obligation to provide this data should have been in the original conditions establishing the various schemes such as through grant agreements or other contracts with the service providers.

Note that this data is different to the performance or operational information required under the *Commonwealth Grant Rules and Guidelines* (CGRG)[31] or the FOI Act.[32] As per the health example in the Interim Report, the data that needs to be made available is that which relates directly to the function of government for which public funds are provided, not (just) the performance of the entity providing the service on behalf of the government.

In the short term, templates for grant guidelines, grant agreements and procurements should be updated to include model clauses[33] to ensure the optimum amount of data is collected and, as a

[25] Digital Economy Strategy
[26] Australian Data Strategy
[27] OECD, Enhanced access to and sharing of data (EASD) - Reconciling Risks and Benefits for Data Re-use across Societies (2019)
[28] Investigating the Link Between Research Data and Impact - ARDC
[29] APCO's Co-Regulatory Model
[30] Product stewardship in Australia - DCCEEW
[31] Commonwealth Grants Rules and Guidelines | Department of Finance, 2017, para 12.8, p.33.
[32] https://www.legislation.gov.au/Details/C2022C00293, s 8A.
[33] ClauseBank | Department of Finance

minimum, accessible by the government body administering the program, service or scheme. It is likely this will require some awareness raising and education as well as monitoring of adoption. Wherever possible, current contracts should also be varied to accommodate this requirement. Lastly, and as a minimum, the metadata of this data should be discoverable for potential sharing via the DAT Scheme.

In the longer term, it may be necessary to amend the CGRC or the FOI Act to clarify the requirement of agencies to ensure data held by government-funded service providers is made FAIR - that is, the data is Findable, Accessible, Interoperable and Reusable for researchers and the public.[34] This will ensure procurement processes satisfy both corporate accountability requirements of government bodies as well as public transparency and researcher productivity through improved availability and use of public data.

Absorptive capacity and the R&D Tax Incentive. The ARDC previously provided a submission to the Australian National Audit Office (ANAO) on their performance audit into the *Administration of the Research and Development Tax Incentive*.[35] In that submission, the ARDC noted issues also relevant to this inquiry affecting additionality and spillover relating to data including the following:

- Under the *Industry Research and Development Act 1986*,[36] AusIndustry (on behalf of Industry Innovation and Science Australia) manages registration of companies accessing the program and determines the eligibility of R&D activities. As part of this process, companies provide information about the Fields of Research (FoR)[37] of the core and supporting R&D activities they have self-assessed as being eligible under the legislation. Of note, the resulting Register of R&D entities and activities is not publicly accessible. The ARDC argues this Register should now be made available based on the same premise as recent changes made for claim data.[38] That is, this data relates to publicly funded R&D rather than the tax system itself (the reason historically given for not sharing the data). If key elements of the Register of R&D entities and their activities were public, it would enable others to identify and therefore potentially collaborate with companies on relevant R&D activities.
- The public Register of Industry R&D should also include metadata for research outputs produced as a result of public funding. This metadata would not necessarily allow access to the contents of the resource itself. The decision on who could gain access would remain with the entity that owned the rights of each resource. Nonetheless, access to the metadata of outputs would itself be highly valuable for a range of stakeholders and purposes.

[34] FAIR Principles
[35] Administration of the Research and Development Tax Incentive | Australian National Audit Office
[36] *Industry Research and Development Act 1986*, Schedule 2.
[37] Australian and New Zealand Standard Research Classification (ANZSRC), 2020
[38] *Treasury Laws Amendment (A Tax Plan for the COVID-19 Economic Recovery) Act 2020*, Schedule 6, Part 1.

As identified in the *Research and Tax Incentive Review*, 'improved access to (R&D Tax Incentive) data by policymakers and the public would facilitate programme evaluation and expose recipients to appropriate public scrutiny around their receipt of public funds'.[39]

For the ARDC and researchers more broadly, it would facilitate insights into research entities, activities and outputs nationally thereby allowing connections between artificially siloed parts of the national research ecosystem. The ARDC is presently working to integrate these data types into a more complete knowledge graph of the entire Australian research system in a project titled 'Research Link Australia'.[40] This capability will be one of the enablers of establishing 'a modern data driven approach (to Excellence in Research Australia)', as expected by the recently appointed Minister for Education.[41]

In the short term, Research Link Australia will make it easier for businesses, policymakers and citizens to 'pull' research expertise, research conclusions and research products. With additional data, such as that suggested above, it will also make it easier for researchers to shape their research projects and 'push' results to those organisations most likely to benefit from their findings.

Access to publicly funded research. There are numerous policy documents to which Australia is a signatory and which encourage adoption of improved access to research from public funding. This includes the *Recommendation concerning Access to Research Data from Public Funding*[42] as well as the UNESCO's *Recommendation on Open Science*.[43] Both of these have five year reporting windows from 2021 after which Australia should report improvements against the stated commitments.

In support of these policies, the ARDC already advocates for and supports adoption nationally of the FAIR Principles for digital research objects.[44] The intention is to make research Findable, Accessible, Interoperable and Reusable for both researchers as well as users of research.[45]

To achieve FAIR, the ARDC implements national digital research infrastructure in ways that recognise:

> …re-use and value of data can depend on the availability of relevant metadata, algorithms, code, and software, from public funding together with information on workflows and the computational environment used to generate published findings, and that providing access to these other research-relevant digital objects from public funding along with the data itself can be essential.[46]

Access to these objects is essential because they support critical pillars of research itself, notably research quality, research integrity, research reproducibility and research impact.

---

[39] Review of the R&D tax incentive, p.23.
[40] New Capability for Linking Industry and Research | ARDC | Australia Research Data Commons
[41] Statement of Expectations 2022 | Australian Research Council
[42] Recommendation of the Council concerning Access to Research Data from Public Funding
[43] UNESCO Recommendation on Open Science
[44] FAIR Principles
[45] The ARDC prefers use of FAIR Principles rather than 'Open' as it recognises the work of researchers can contain sensitivities preventing all objects from being openly available. FAIR and Open are overlapping and complementary concepts.
[46] Recommendation of the Council concerning Access to Research Data from Public Funding

In terms of this inquiry, the benefits of this approach are that:

- The cost of research is reduced by increasing the availability of inputs and decreasing the cost to access them as well as by decreasing follow-on costs for users accessing research outputs.
- A more representative array of research outputs is formally recognised thereby reflecting the actual productivity of Australian researchers in terms of the value they create.[47]

In the short term, all governments should mandate the immediate publishing of at least the metadata of research outputs to ensure they are appropriately Findable regardless of jurisdiction, sector or program. Importantly, the definition of research outputs must be broadened (in line with international trends) beyond traditional research outputs so as to include not only journal articles and books, but also data, software, vocabularies, and more. This should apply 'to all publicly funded R&D' as defined by the OECD's 'Government budget allocations for R&D' (GBARD).[48] GBARD is already used by the Australian Bureau of Statistics (ABS)[49] and the Department of Industry and Science (DISR)[50] for statistical reporting of R&D nationally.

Use of Persistent Identifiers. Use of persistent identifiers (e.g. ORCID[51] and DOI[52]) across publicly funded research grant programs remains ambiguous despite the infrastructure, services and support being available for many years. If mentioned in grant program guidelines at all, identifiers are usually 'optional' or else 'must be supplied if known'.

Use of globally unique and persistent identifiers should be mandatory for key components across the national research ecosystem. This extends beyond academia and needs to include government and industry research. If adopted more fully, this would ensure traversable links between budgets funders,[53] grants,[54] institutions,[55] research projects,[56] research outputs[57] (including data[58]) and citation analysis.

In support of greater adoption of this approach, a recently completed cost-benefit analysis revealed that using persistent identifiers in the Australian research sector could save $24 million per year and 38,000 person days in wasted effort every year.[59] Primarily, this wasted effort is a result of unnecessary re-keying

[47] A roadmap toward a common framework for measuring the Digital Economy | OECD, para 4.2, p.54.
[48] Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development.
[49] Research and Experimental Development, Government and Private Non-Profit Organisations, Australia, 2020-21 financial year | Australian Bureau of Statistics
[50] Science, research and innovation (SRI) budget tables
[51] ORCID
[52] DOI
[53] https://www.crossref.org/services/funder-registry/
[54] https://www.crossref.org/blog/global-persistent-identifiers-for-grants-awards-and-facilities/
[55] https://ror.org/
[56] https://www.raid.org.au/
[57] https://www.doi.org/
[58] https://datacite.org/
[59] Strategic Investment in Identifiers Could Save $24 Million and 38,000 Person Days per Year | ARDC

of information by researchers across multiple systems. However, if adopted broadly, other benefits include better national coordination of research funding across jurisdictions and across programs.

In terms of the use of grant identifiers, GrantConnect[60] does represent an important initiative by the Commonwealth's Department of Finance including through use of identifiers, however, the service is not without its limitations.[61] In the view of the ARDC, this program should have:

- delivered public APIs for GrantConnect and ensured data was discoverable via data.gov.au.[62]
- provided grants data publicly as 'linked data' utilising various data services already available nationally as well as building and publishing new ones as required.[63]
- extended reporting to include deliverables or outputs from grant funding where relevant to better support measurement of return on investment from grants programs nationally.
- developed a roadmap to capture grants nationally with the cooperation of states and territories.

Some of these shortfalls are indicative of a broader trend of presenting public sector data only through portals for use by the general public rather than also enabling machine-to-machine interactions for sophisticated users. While investments might be approved with this in mind, it appears not to translate through to design and implementation of services as per the *Digital Government Strategy*.[64] This hints at a structural lack of coordination or architectural control across programs. Without this, it will be difficult to achieve the benefits as promised from adoption of a 'digital first' approach by the government.

Improved Access to Standards

The ARDC supports the intention to improve access to Standards in line with Recommendation 3.3 of Interim Report 3.[65] Standards based development of digital ecosystems both within Australia and internationally is critical to attaining levels of assurance and interoperability necessary to improve productivity. As such, the OECD recommends, Adherents (that includes Australia) 'should support the development, maintenance, adoption, dissemination, and implementation of technical standards that are open, freely accessible, and internationally agreed to the greatest possible extent.'[66]

Australian governments effectively create national or state standards for data encoding through a range of data collection or aggregation programs. For example, the standards implemented by the ABS in the census or by the Australian Institute of Health and Welfare (AIHW) in national health data establish standard classifications and data terminologies that then inform research, administration, and private sector activity well beyond the original scope or purpose of the administrative standard classification.

[60] GrantConnect
[61] Operation of Grants Hubs | Australian National Audit Office
[62] Application Programming Interface. Refer https://en.wikipedia.org/wiki/API for further information.
[63] Australian Government Linked Data Working Group
[64] Digital Government Strategy, p.17.
[65] Interim Report 3, Recommendation 3.3, p. 80.
[66] Recommendation of the Council concerning Access to Research Data from Public Funding

This is a positive contribution to semantic interoperability across the nation, however government agencies are seldom resourced to disseminate, support and update these standard terminologies, such as in data vocabulary services.[67] This national information infrastructure is often an afterthought or mere by-product of administrative programs rather than a critical resource that agencies sustain as a public good benefitting the broader economy and society.

As an added benefit these government data standards are often themselves aligned to OECD or UN reporting programs and thus lay the foundation for international scientific or industrial collaboration.

Governance and Funding Models

It is routine for government entities to receive and expend funds to fulfil their functions. Once funds are received, the entity procures various inputs from suppliers, assembles those inputs in accordance with proposed designs, and delivers the outputs to stakeholders with the intent of achieving outcomes. The department's corporate governance ensures it is all done efficiently and effectively.

The challenge of national data and digital ecosystems is that they disrupt this linear model. Perplexingly, data sharing programs can be thwarted if they try to 'buy' data from suppliers, 'assemble' it centrally and deliver it to stakeholders (who are often the same entities that supplied the data).

The characteristics of data negates the ability of the Commonwealth to be an indispensable intermediary through which all data must pass before it is considered a 'national data asset'. Instead, a national data asset emerges from data providers independently publishing their data to a collectively agreed standard. In this context, the Commonwealth does not internalise production of the data asset or indeed even have to be present. However, the Commonwealth does have valuable data itself to contribute and is also ideally placed both to orchestrate disparate players and sponsor the provision of 'federated services'.[68]

So, while the key challenge is in enabling, incentivising and sustaining data supplies, it is often the data aggregator or users (rather than the data suppliers) that are given the resources and responsibility to make it happen. That is, a Commonwealth body is funded to deliver a (centralised) national data asset within a multi-year timeframe, but is reliant on data from states and territories whose budget cycles, resources and priorities are not always able to be aligned prior (or even during) the program. Additionally, departments are reluctant to fund jurisdictions directly to provide data as it risks monetising a public good, and it risks data supplies drying up after program funding ceases.

To further confound the issue, departments typically fill program boards with only their key staff (Chief Financial Officer, Chief Information Officer, etc) as a way of satisfying Commonwealth 'corporate governance' responsibilities for program outcomes (as interpreted by them). Data suppliers can typically

[67] About - Research Vocabularies Australia

[68] 'Federated services' are operated by trusted third-parties to facilitate the trusted peer-to-peer exchange of data between participants in a data sharing ecosystem. Examples of federated services include data catalogues, vocabulary services as well as authentication and authorisation services.

be relegated only to semi-regular participants of working groups or advisory committees often with minimal or no decision rights, particularly regarding funding.

This tension between vertical 'corporate governance' and the need for improved horizontal 'public governance' or 'public sector governance' is not new.[69] However, the necessity for better collaborative or networked governance is arguably more salient for digital ecosystems. This is because achieving the outcome relies on the actions of legally autonomous suppliers who will only share data once they trust the sharing mechanisms, and have the resources necessary to transform and publish their data.

Australia does, of course, already have very many joint ventures and cooperatives, including for data sharing purposes, so clearly the legal structures exist. Some domestic examples include AgRefed,[70] Geoscape Australia[71] and Austroads.[72] Internationally, there are model research consortia agreements[73] as well as model data collaborations for industry[74] as well as more generally.[75] These all rely on an archetype of successful data sharing ecosystems with characteristics that include:

- An administering organisation trusted to act in the best interests of the collective
- A horizontal 'collaborative' governance model over a vertical 'corporate' model[76]
- A de-centralised funding allocation over which the collective has substantive decision rights
- A program that is requirements driven rather than schedule driven[77]

If Australia is to improve data sharing nationally and thereby achieve the intended productivity dividends, then the need for better collaborative governance and funding models for national data and digital ecosystems has to be refined further. This includes providing practical guidance and support to practitioners tailored to the peculiarities of data and digital ecosystems. In this scenario, the Commonwealth has the opportunity to be a 'crucial catalyst for essential reform'.[78]

Data Sharing and Integration Opportunities

Interoperability of data and digital ecosystems is a key enabler of productivity. Interoperability must go beyond technical interoperability (the ability for machines to message each other) and extend to include semantic interoperability (the ability to understand the meaning of data that is exchanged) and

[69] Public Sector Governance in Australia, Ch 1. Refer to 'Framing the taxonomy of governance'.
[70] AgReFed
[71] Geoscape Australia
[72] Austroads
[73] DESCA Model Consortium Agreement
[74] Design Principles for Data Spaces
[75] Rulebook for a fair data economy - Sitra
[76] Modes of Network Governance: Structure, Management, and Effectiveness | Journal of Public Administration Research and Theory | Oxford Academic
[77] The Influence of Project Structure and Governance on Systems Engineering and Project Management Relationships - SEBoK
[78] Shifting the Dial: 5 year productivity review - Inquiry report, p.8.

pragmatic interoperability (the practical ability to interoperate, such as appropriate access and use of data across systems).

In Australia, there are already a large and growing number of data sharing initiatives, many focussed on enhancing the work of users in specific domains or organisations. These are important pathfinder and foundational activities, but the next phase will need to be characterised by a shift in emphasis ensuring local systems integrate more fully with systems nationally and internationally. This is not only important from the perspective of cost and user experience, but is necessary to make the most of the growing number of smart services (such as Artificial Intelligence and Machine Learning) with a key outcome being improved productivity gains nationally.

To take an example from current initiatives, Interim Report 2 notes[79] the Action Plan of the *Australian Data Strategy*[80] mention 'sector-specific initiatives in areas such as collecting and sharing freight data,[81] modernising waste data visualisation[82] (and) integrating regional datasets….[83]' Each of these are important, but what is not clear is the frameworks for how these initiatives integrate with each other (crucially, at all the interoperability levels mentioned previously) as well as with similar research (and industry) sector instances.[84] If this interoperability were made possible, it would enable more complex insights for less effort, such as smart services enabling stakeholders to ask and get answers to questions such as, 'What are the critical waste transport issues affecting regional Australia?'.

The Australian government is not alone in facing this stage of development. Similarly, the ARDC is undertaking work called 'Thematic Research Data Commons' to scale-up digital research infrastructure to meet Australia's future research needs.[85]

However, the benefits of this effort by the ARDC (and equivalent effort by others) will be limited if this is not coordinated more broadly across sectors nationally. As an example of how this is being done elsewhere, the *European Data Strategy*[86] supports nine 'data spaces'[87] based on strategic economic sectors and domains of public interest. These include: Industry (Manufacturing), Environment, Mobility, Health, Finance, Energy, Agriculture, Public Administration and Skills. Critically, the Strategy conceives these nine data spaces as able to operate as a single coherent European data space. That is, every data

[79] Interim Report 2 - 5 Year Productivity Inquiry: Australia's data and digital dividend, Box 3.1, p.43.
[80] Australian Data Strategy Action Plan
[81] National Freight Data Hub
[82] National waste and resource recovery data hub - prototype website - DCCEEW
[83] Regional Data Hub: supporting delivery in our regions | Department of Infrastructure, Transport, Regional Development, Communications and the Arts
[84] The Australian Transport Research Cloud | ARDC
[85] Designed for the Future — Thematic Research Data Commons | ARDC
[86] 2020 European Data Strategy
[87] Data spaces - Gaia-X: A Federated Secure Data Infrastructure

space must be interoperable with all other data spaces. And this interoperability must encompass business, legal, security, technical and semantic themes.[88]

It is also planned that each data space can access a Data Innovation Hub (an equivalent data space tailored for research) to facilitate applied research and experimental development with industry. In this sense, Data Innovation Hubs are analogous in function to the ARDC's thematic research data commons that must also be interoperable with at least their respective government and industry counterparts.[89]

It is worth noting the ARDC has not yet finalised the partitioning of domains for thematic research data commons. That is, ARDC thematic research data commons may not align with the nine data spaces of the European Data Strategy; ideally, focal areas here would be coordinated by the Commonwealth (with an eye to interoperability globally). Initially, the ARDC has started with two thematic research data commons - the 'People Research Data Commons'[90] and the 'Planet Research Data Commons'.[91] These are pilots within domains familiar to the ARDC and supported by detailed analysis regarding their priority.

Of note for this inquiry, a key aim of the thematic data commons is achieving an intent similar to that of the European Data Innovation Hubs relative to data spaces. That is, better facilitating applied research and experimental development with industry and government. This can be challenging as research systems and operational systems are usually separate and typically optimised for their respective roles. While willing to cooperate, neither benefit from moving to a single system. As such, the thematic research data commons aims to dovetail researchers and practitioners together in a way that translation can occur more easily and sustainably.

For the vision of improved data and digital interoperability to come to fruition in Australia, the government must take a lead role in sponsoring adoption of shared design standards and approaches nationally similar to the data spaces concept.[92] That is, both within and between different sectors. Importantly, if the objectives of the *Australian Data Strategy* and *Australia's Digital Economy Strategy* are to be realised, the government will need to lift efforts significantly beyond technical interoperability as well as focus outwards more, beyond just the government sector or just within stakeholders of each government portfolio. This is because the challenges we face, and therefore the data we need, clearly cut across the structures that we choose to use when organising ourselves.

Should you wish to discuss these or other matters, please contact Dr Adrian Burton, Director Data Policy & Publication Services

[88] Data Interoperability Maturity Model | naa.gov.au
[89] Note that the ARDC has not yet finalised the partitioning of the domains for each Thematic Research Data Commons. That is, our Thematic Research Data Commons may not align with the nine data spaces of the European Data Strategy.
[90] People Research Data Commons | ARDC
[91] Planet Research Data Commons | ARDC
[92] International Data Spaces