Submission: Draft Data Availability and Transparency Code 2022

Rosie Hicks, CEO ARDC

26/09/2022

The Australian Research Data Commons (ARDC) thanks the Office of the National Data Commissioner (ONDC) for the opportunity to comment on the draft *Data Availability and Transparency Code 2022*.[1]

About the 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.

Introduction

The Consultation Paper states that, 'the Data Availability and Transparency Act 2022 (the Act) establishes a new, best practice scheme for sharing Australian Government data – the DATA Scheme', and that 'the DATA Scheme is underpinned by strong safeguards and consistent, efficient processes'.[2] The ARDC provides this submission in the spirit of collaborating with the ONDC to ensure the Scheme does, in fact, reflect best practice.

A key theme of this submission is that many provisions within the draft Code should be embedded within the Accreditation Framework rather than bilateral sharing arrangements. Essentially, the Accreditation Framework should be the key trust mechanism thereby reducing the overall cost of every subsequent data exchange.

If this approach is not adopted, data sharing agreements are likely to have issues similar to Memorandums of Understanding (MoU). In 2016, the Department of Prime Minister and Cabinet noted MoU '...can be unnecessarily complicated and time consuming….tak(ing) several years and multiple memorandums to establish data sharing arrangements between…entities….incurring significant cost in policy, project and legal officer time'.[3]

[1] Consultation on the Data Code
[2] ibid, p.3.
[3] Guidance on Data Sharing for Australian Government Entities

ARDC is enabled by NCRIS

Another theme is the desire to leverage existing mechanisms of the university sector that align with intended controls of the DATA Scheme, such as public interest tests. This will ensure consistent and efficient processes.

ARDC Response to Consultation Questions

Project principle: project reasonably expected to serve the public interest

1. ***Is the approach to weigh arguments for and against the project serving the public interest appropriate? If not, how else could entities assess whether a project for the purpose of informing government policy and programs, or research and development, serves the public interest?***

The ARDC is concerned this section enables custodians to re-evaluate the public interest test of publicly funded research projects that have already met this requirement variously as part of national research codes, funder research policies, grant guidelines, peer review processes, due diligence, ethics approvals,[4] Ministerial approvals or Cabinet approved policies. For example, the current draft would see custodians re-evaluating an explicit National Interest Test statement already approved by a Minister as part of Australian Research Council funding processes of the National Competitive Grants Program (NCGP).[5]

To achieve the intended consistency and efficiency of the Scheme, it is recommended the Code should state that, 'if the only data sharing purpose of the project is research and development (see s15(1)(c) of the Act) and it is funded publicly by an Australian government, the project can reasonably be expected to serve the public interest'.

The definition of 'publicly funded research and development' should align with the OECD's 'Government budget allocations for R&D'[6] as already used by both the Australian Bureau of Statistics[7] and the Department of Industry, Science and Resources.[8]

2. ***If yes to the above are the requirements of what entities must do, to weigh up arguments for and against the project serving the public interest, clear and unambiguous, and is this list proper and pragmatic? In your response, please provide reasons.***

3. N/A

4. ***Is the list of projects that do not serve the public interest able to be practically applied? What, if any, further guidance is required to support entities when a project does not serve the public interest?***

As per Question One above, publicly funded research and development projects under the DATA Scheme should be treated as for the purpose of 'delivery of government services'. That is, not subject to these additional evaluation criteria. These criteria could be used for research and development projects not funded publicly.

[4] For example, the specific requirement for Human Research Ethics Committee to evaluate public interest under Section 95A of the Privacy Act 1988. Guidelines approved under Section 95A of the Privacy Act 1988 | NHMRC
[5] Articulating National Interest in grant applications | Australian Research Council
[6] Organisation for Economic Co-Operation and Development's Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development.
[7] Research and Experimental Development, Higher Education Organisations, Australia, 2020 | Australian Bureau of Statistics
[8] Science, Research and Innovation (SRI) Budget Tables.

5. ***Are the notes contained in this section helpful, and would this section benefit from other illustrative examples provided as notes? If yes, what examples and under which subsections?***

Nil response.

Project principle: applicable processes relating to ethics

6. ***Under the draft data code, entities must have regard to any process of ethics applicable. Do you have any comments about this approach?***

The ARDC welcomes the approach to leverage one or more existing ethics approvals processes as agreed by both parties. It may be beneficial to make it clear that an ethics approval process is not always required. Additionally, should the custodian require one or more ethics approval processes where none is otherwise required by the data user, any costs of the additional processes must be borne by the custodian.

It is hoped that the Scheme will in future catalyse streamlining of ethics approvals by enabling states and territories to coordinate and reduce duplication or even multiplication of ethics approvals required for national or cross jurisdictional data assets. In this way, the DATA Scheme could complement and build upon progress already achieved under the *National Certification Scheme for the ethics review of multi-centre research*.[9]

7. ***Is the note provided to assist entities identify ethics processes helpful? Why, or why not?***

Nil response.

People principle: conflicts of interest

8. ***Are the requirements of this element of the people principle clear and unambiguous? What, if any, further details or guidance could assist?***

While this section is clear it would be better if, rather than implying credentials are to be re-evaluated for every data sharing agreement, robust processes of Collectors were instead assured by the Accreditation Framework. That is, accreditation assures custodians that Collectors have a demonstrated capacity to comply with ss 16(3) and 16(4) for any agreement entered into as part of the DATA Scheme.

9. ***Is the example provided under this section helpful? Why, or why not?***

Nil response.

People principle: appropriate persons

10. ***Are the attributes, qualifications and affiliations listed in this section appropriate and easy to understand?***

The ARDC notes that s16(a) of the Act refers to ('designated', as in classes of)[10] individuals requiring them to have the 'attributes, qualifications, affiliations or expertise appropriate for the access'. Section 16(b) (in which the Commissioner has latitude to make 'any other matters specified in a data code')[11] refers only to the attributes of 'the Collector'. The draft data Code appears to conflate these two provisions. This seems to result in an

[9] National Certification Scheme for the ethics review of multi-centre research | NHMRC
[10] s123
[11] s16(4)(b)(iv)

undesirable extension of the capacity of the Sharer to decide access rights down to (potentially) 'named' individuals. This level of discretion should instead remain with the Collector in accordance with their obligations (and demonstrated capacity as per the Accreditation Framework) as an entity under the DATA Scheme.

In the view of the ARDC, this section should be revised to avoid Sharers having the ability to be overly prescriptive about the individuals who can access data and instead cover only those 'matters relating to the (capacity of the) entity collecting the data' as the means of ensuring access is by appropriate persons.[12] This does not remove the ability of custodians to describe the attributes of those who can access the data, such as 'Australian citizen'.

11. ***Would this section of the draft data code benefit from other illustrative examples provided as a note? If yes, what examples and under which subsections?***

Nil response.

Setting principle: reasonable security standards

12. ***Is this section adequate in clarifying what are reasonable standards?***

The ARDC is concerned that data custodians could use this section to frustrate or delay the sharing of data by imposing unnecessarily detailed security requirements that are bespoke to every data sharing agreement.

It would be preferable if security levels were defined within the Accreditation Framework and conformance with them assured by an independent third party assessor registered with or recognised by the ONDC.

In this scenario, the Sharer of data would define the level or standard of data security required for each dataset in accordance with the accreditation standard, and only those users or systems independently accredited to that level are eligible to consider receiving or handling that data.

13. ***Would this section benefit from an illustrative example provided as a note? If yes, what are some proposed examples?***

Nil response.

Data principle: appropriate protection – whether data should be altered

14. ***In practice, this element of the data principle, the privacy protections, and three data services set out in the Act, all work together to provide a framework to appropriately protect data. ONDC acknowledges there is a need to strike the right balance between taking a layered approach and not making the DATA Scheme too complex. Could the draft data code be improved to better assist entities apply this element of the data principle?***

The ARDC is supportive of the principle of data minimisation. To facilitate this approach, the data Code should stipulate that custodians must publish certain information in addition to the metadata of their datasets. In particular, the data Code should require custodians to provide links to data models alongside data descriptions in data catalogues. The details of what is required could be covered in Guides issued by the Commissioner. If this information were made available by custodians, users would then be able to request only the minimum data elements or data points necessary to achieve the intended outcomes of the research and development project.

[12] ibid.

Lastly, s12(4) of the data Code should be reviewed to ensure it does not inadvertently prevent or deter custodians from using Privacy Enhancing Technologies, such as federated learning or homomorphic encryption.[13] For example, control over the data may not require 'removing or altering the data', but could instead be achieved by controlling the operations that algorithms can perform in relation to data that remains with the custodian. These approaches might be applied by custodians in collaboration with one or more users and not just by an ADSP.

Data principle: appropriate protection - data sharing must be reasonably necessary

15. ***Is the 'reasonable person' test adequate in this section? If not, how could this section be improved to allow the entities to test whether the data proposed to be shared, collected and used is reasonably necessary to achieve the data sharing purpose?***

Nil response.

Output principle

16. ***In practice, the output principle requires entities to agree how the accredited user will use shared data. Overall, how could the draft data code be improved to best assist entities apply the output principle?***

Nil response.

Privacy protections

17. ***One of the objects of the Act is to enable the sharing of data consistently with the Privacy Act and appropriate safeguards. Does this part of the draft data code strike the right balance between holding data custodians accountable to seek consent, and providing data custodians with an exception to collect consent in circumstances where it is genuinely unreasonable or impracticable to seek consent? How could the draft data code be improved to achieve the right balance? For example, could the National Health and Medical Research Council waiver of consent guidelines be used here?***

Nil response.

18. ***Is this part of the draft data code adequate in providing further clarification for what considerations should be taken into account when determining whether it is necessary to share personal information to properly deliver a government service? How could this section be improved?***

Nil response.

19. ***Does this part of the draft data code provide an adequate list of factors for data custodians to consider when determining whether the public interest justifies the sharing of personal information without consent? Would this section benefit from an example provided in a note, and if so, can you suggest one?***

Nil response.

[13] Privacy-enhancing technologies - Wikipedia

Data sharing agreements

20. ***Should the data sharing agreement include any additional details about the designated individual who is a foreign national?***

The ARDC considers this section as an unnecessary obligation on universities, particularly in context of other provisions of this and other Acts. As such, it should be revised or preferably removed entirely.

In terms of access for research and development, the Act already limits participation in the DATA Scheme to 'Australian Universities' 'established by or under a law of the Commonwealth, a State or a Territory'.[14] This, for example, precludes a number of overseas universities already operating within Australia. Access is also restricted to 'designated individuals' within those Australian universities.[15] Furthermore, the General Privacy Provisions in the Act at s16A(2) state that:

> *If data that includes personal information is shared, the data sharing agreement that covers the sharing must prohibit any accredited entity …* ***from storing or accessing, or providing access to, the ADSP-enhanced data, or the output, of the project outside Australia****.(ARDC bold)*

In light of these provisions as well as other regulatory considerations (e.g. privacy, cybersecurity, critical infrastructure, foreign arrangements and foreign interference) it is not clear this provision of the draft Code is necessary or appropriate, particularly for the vast array of data that does not include personal information.

This provision also results in the collection of data on a large number of individuals that may not be relevant, and it is not clear universities (or indeed individuals) will always be aware of foreign nationality (as evidenced by the recent experience of parliamentarians). As such, the practicality of complying with this provision is questionable.

Lastly, it is also not clear that the data sharing agreement is the most appropriate mechanism to capture this information. First, the data sharing agreement is unstructured or semi-structured and therefore it will be difficult to query at scale. Second, data sharing agreements should presumably be completed prior to sharing. While universities will know the 'designated individuals' able to access the shared data, they may not know or may want to change over time the actual individuals involved. This would result in universities having to make frequent amendments to data sharing agreements already registered with the ONDC by data custodians.

It would be preferable if instead the Accreditation Framework ensured universities kept immutable audit logs fully identifying those who accessed shared data. These logs could, if required and after appropriate approvals were granted, be queried by the ONDC or data custodians indicating when and by whom each dataset was accessed. This data could be combined by the ONDC with immigration or other datasets to determine the nationality or visa status of users (if this data is not already shareable by universities through their identity and access management systems). The details of the data query point for these audit logs and necessary approval processes could be included as additional information to data sharing agreements as described in this draft data Code at s21(1)(c).

[14] s9
[15] s123

Miscellaneous

21. ***This part of the draft data code is informed by the list prescribed in section 130 of the Act. Is this an appropriate approach, and are there any additional details that should be provided to the Commissioner outside of that list?***

Nil response.

22. ***Is 31 July an appropriate deadline for data custodians to provide information and assistance to the Commissioner to prepare for the annual report?***

Nil response.

Potential additions to the data code

23. ***What additional topics could the data code include to assist the establishment or integrity of the DATA Scheme?***

As described in a previous ARDC submission, the ONDC should provide a range of 'interoperability services' that facilitate the trusted sharing of data between entities.[16] Interoperability services operated by the ONDC to support the data sharing principles should be described in this data Code. The Dataplace function might be one example.[17]

These interoperability services are essential for facilitating near real-time sharing or controlled access consistent with terms stipulated by data custodians. Importantly, these interoperability services are not bodies through which data must pass or else managed repositories in which data must be stored, but are instead services that apply controls over any involvement of entities with data. One analogy is the automated service a bank uses to ensure that the PIN entered is a match for the card used whenever a customer wants to withdraw cash from a teller machine. Of note:

- Some of these services will be automated, such as authentication, authorisation and auditing of transactions, but others will rely on manual processes, such as accrediting entities under the Accreditation Framework.
- Some of these services might be from established providers who may be accredited to operate under the DATA Scheme should they choose to do so, such as the Australian Access Federation; others may be newly established by the Commissioner solely for the purpose of supporting the DATA Scheme.

Regardless, use of these interoperability services should be mandatory to ensure predictable levels of trust and control by all parties when making use of the DATA Scheme to share data between participating entities.

Should you wish to discuss these or other matters, please contact [redacted]

[16] 20201106_Submission on the ONDC DAT Bill Accreditation Framework_ARDC
[17] Use Dataplace | Office of the National Data Commissioner