Data Sharing and Release: Legislative Reforms Discussion Paper

Department of Prime Minister and Cabinet

22 October 2019

Executive Summary

The University of Melbourne welcomes the opportunity to respond to the Department of Prime Minister and Cabinet's *Data Sharing and Release Discussion Paper*.

As a leading research institution, we have a particular interest in positive legislative reforms in this area. Data sharing arrangements that ensure trusted users have access, for designated purposes and in safe environments, to quality public sector data is crucial to the performance of Australia's research sector and to the public impact that it delivers. It is important that the Government seek to better enable safe data sharing, from a truly national perspective, in view of the improved research and public policy outcomes that this promises.

The direction for legislative reform set out in the Discussion Paper is overwhelmingly positive. The Paper rightly recognises that data integrity and managing privacy-related risks is of paramount importance when dealing with public sector data. It also recognises that the way to achieve this is by ensuring that data is shared in a responsible or safe way, rather than by being inhibited by poor data practices or by preventing access altogether. The proposed legislative reforms include a suite of sensible measures that support integrity in data sharing, including limits on the purposes for which data can be shared, safeguards for dealing with privacy risks, and transparency and accountability mechanisms such as public registers for Data Sharing Agreements and for Accredited Data Service Providers and Users.

The University of Melbourne is broadly supportive of the proposed approach to developing the data sharing legislative framework, and applauds the recognition, in the Discussion Paper, that the task of achieving the right balance in these arrangements requires ongoing attention. We appreciate the progress made so far, and we also believe that there is further work to be done in improving arrangements for the responsible sharing of public sector data with specified users. Our responses to the consultation questions contained below suggest additional considerations that build upon the approach already developed and outlined in the Discussion Paper. These include strongly affirming and potentially elaborating the role of the National Data Commissioner in helping to build and maintain public confidence in the data sharing framework, and the use of differential privacy as a framework for guaranteeing the privacy of individuals in a sensitive dataset. We also note parts of the Discussion Paper where the proposals require further elaboration to inform a proper assessment of them.

We would welcome the opportunity to further discuss future developments in the data sharing legislative framework as the development process proceeds towards legislation.

For further information or to discuss this submission please contact Professor Liz Sonenberg, Pro Vice-Chancellor (Digital and Data)) at [redacted]

Discussion questions

1. Do you think the distinction between data sharing and data release is clear? How could this distinction be clearer?

The Discussion Paper distinguishes between data sharing and data release. However, the use of the phrase "*open* data release" is potentially confusing.[1] Since released data is open by definition, the adjective "open" is unnecessary and therefore should be omitted.

2. What are the challenges for open release of public sector data?

Two key challenges for the release of public sector data are: the uneven capability in relevant departments and agencies in relation to data practices for management and sharing of data; and establishing adequate safeguards against re-identification. The release of the Medicare Benefits Scheme/Pharmaceutical Benefits Scheme dataset,[2] as well as the intended release of the Victorian Myki dataset via open data channels,[3] point to the challenges related to privacy protection, with analysis of datasets showing it possible to re-identify individuals in both cases. It is crucial that measures are put in place to ensure that data released to the public do not allow individuals to be re-identified in this way. For example, differential privacy is a relatively new framework for guaranteeing the privacy of individuals in a sensitive dataset, when releasing aggregate statistics or machine-learned models on such data and is already in use by the U.S. Census Bureau[4]. This should be part of the approach to privacy-by-design in data sharing systems. The Discussion Paper does not consider this. Further, consideration should be given to the potential value of consistency with the Privacy Act that distinguishes between primary and secondary uses, and the possibility of separate hurdle requirements.

Issues relating to Indigenous data sovereignty represent a further set of challenges for legislative reform of arrangements for data release. The University endorses the considered and cautious approach that the Commission is taking to addressing the opportunities and risks that the legislation could present for Aboriginal and Torres Strait Islander peoples.

3. Do you think the Data Sharing and Release legislative framework will achieve more streamlined and safer data sharing?

The Data Sharing and Release Legislation represents an opportunity to promote the benefits generated through sharing data with specified users while also ensuring that privacy and data integrity are appropriately safeguarded. If properly designed and implemented, the legislative framework can advance both aims, but we recognise this will be a complex balancing task. Privacy protecting released data is not a solved problem, and a cautious approach to both sharing in controlled environments, and release, is important now, to reduce the likelihood of erosion of social license for legitimate sharing.

It is important to consider ways in which the intended streamlining of data sharing may be undermined by arrangements that are unworkable in practice. We note the points made in The

[1] See, for example, *Discussion Paper*, p.3.
[2] Chris Culnane, Benjamin I. P. Rubinstein and Vanessa Teague. Health Data in an Open World. Dec. 2017. arXiv: 1712.05627 [cs.CY]. https://arxiv.org/pdf/1712.05627
[3] Chris Culnane, Benjamin I. P. Rubinstein and Vanessa Teague. Stop the Open Data Bus, We Want to Get Off. Aug. 2019. arXiv:1908.05004 [cs.CR]. https://arxiv.org/pdf/1908.05004
[4] John M. Abowd, The US Census Bureau adopts differential privacy. Proceedings of the 24th ACM SIGKDD Internaitonal Conference on Knowledge Discovery and Data Mining. 2018.

Australian & New Zealand Real-World Data Network (RADiANT) response to the Discussion Paper concerning the potentially onerous burden placed on researchers and data custodians by the proposed process for data sharing. The process may be particularly burdensome when dealing with integrated datasets drawn from multiple sources (e.g. National Interest Datasets): the process proposed in the Discussion Paper seemingly requires a data sharing agreement with the primary custodians from each data source in such cases. This issue could be addressed by establishing a single body to manage data sharing agreements concerning integrated datasets.

4. What do you think about the name, Data Sharing and Release Act?

The University of Melbourne suggests renaming the Act to *Responsible Data Sharing and Safe Release Act*. Adding "responsible" and "safe" to the name will highlight the steps being taken to ensure data is only being shared under conditions that safeguard against privacy breaches and against misuse.

There is also a question as to whether the Act's name should include the term "Release" at all, given that the focus of the legislative framework is on data sharing rather than data release.

5. Do the purposes for sharing data meet your expectations? What about precluded purposes?

The University of Melbourne endorses the three data sharing purposes articulated in the Discussion Paper: research and development, policy and programs and service delivery. These purposes capture the broad benefits that will be promoted through improved arrangements for the sharing of public data with specified users.

While the purposes identified are largely appropriate, there is scope for further refinement. It may be useful if the purposes were more clearly defined than they have been in the Discussion Paper. For example, some research and development has a public benefit as its specific aim e.g. progress in understanding a given public policy area. In other cases, the relevant public benefits may form only a secondary motivation or an unintended spill-over benefit. There is a need for clarity on how the new legislation will apply to research programs whose core aims differ in this way. Additionally, there will be increasing opportunities for valuable, controlled access by researchers to data from multiple public sector sources; accordingly the purpose test should be framed as not to inhibit such directions.

We also commend the Discussion Paper for making clear the purposes that are *not* authorised under the proposed legislation. The legislative framework should not enable assurance and compliance activities, nor national security and law enforcement. While of course those represent legitimate government functions, each of these purposes are addressed in other legislation; it is not appropriate that data sharing legislation be used to advance these. Emphasising the limited scope of the new legislation, as well as the safeguards to be put in place to manage the risks associated with data sharing, will help build public support for the legislative framework.

6. What are your expectations for commercial uses? Do we need to preclude a purpose, or do the Data Sharing Principles and existing legislative protections work?

The University of Melbourne does not support an explicit preclusion of data sharing for commercial purposes, but urges particular caution in relation to the possibility that unit-record level data be shared with commercial entities. The commercial use of public data can, if appropriately limited, deliver significant public benefits. For example, research conducted by the Health Research Authority in England found that while people are less accepting of private health providers having

access to anonymised patient data, their attitude became more accepting upon learning of the benefits in terms of improved healthcare products and services.[5] Rather than precluding data sharing for commercial ends, the legislative framework should seek to limit use of the relevant shared data, for example, so that Data Sharing Agreements do not permit uses of data that extend beyond the initial proposal that was deemed within the scope of the identified purposes.

A further issue concerns data collection that is enabled by public funding, but where the data are collected by non-government entities; e.g. where a government department contracts out data collection and/or storage to a third party. The University of Melbourne suggests that the same principles that apply to the sharing of Government-held data should apply to the sharing of data collected and stored by such third parties.

7. Do you think the Data Sharing Principles acknowledge and treat risks appropriately? When could they fall short?

8. Is the Best Practice Guide to Applying Data Sharing Principles helpful? Are there areas where the guidance could be improved?

9. Do the safeguards address key privacy risks?

The University of Melbourne broadly supports the treatment of risk entailed in the Data Sharing Principles, but notes they are not a risk-based framework. A key aim in the legislation should be to strike a balance between appropriately managing the risks associated with data sharing on the one hand, and on ensuring that the data shared is useful to the researchers who access it and therefore promotes the intended public benefits. Poor quality data, and data provision in forms that limit its utility, inevitably undermines policy and research outcomes: this too should be recognised as a risk to be managed in the legislative framework and associated Guides. Risk frameworks such as the OWASP Risk Rating Methodology[6] could usefully be drawn upon in subsequent phases of refinement of the Principles and Guides, as should be international developments in privacy techniques such as differential privacy.

10. Are the core principles guiding the development of accreditation criteria comprehensive? How else could we improve and make them fit for the future?

In broad terms, the three core principles for guiding the development of accreditation criteria are appropriate. However, the adequacy of these principles will depend on how they are defined and operationalised in an accreditation framework. While the first principle, "skills and capability to protect, manage and use data" should form part of the basis for accreditation, there remains a question of how the relevant skills and capabilities are to be delivered. This is pertinent in the context of the present environment which has failed to deliver such capabilities. Regarding the second principle, currently framed in terms of "personal information", we suggest consideration be given to coverage of data about, or derived from, individuals: the form of the data, whether it is unit-record or aggregate, should not matter while the risk to individual privacy remains.

[5] Chico, Victoria, Amanda Hunn and Mark Taylor (2019), "Public views on sharing anonymised patient-level data where there is a mixed public and private benefit" (Health Research Authority and University of Sheffield School of Law).
https://www.hra.nhs.uk/about-us/news-updates/sharing-anonymised-patient-level-data-where-there-mixed-public-and-private-benefit-new-report/

[6] https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology

11. Are there adequate transparency and accountability mechanisms built into the framework, including Data Sharing Agreements, public registers and National Data Commissioner review and reporting requirements?

Transparency and accountability are crucial to building and maintaining public confidence in the data sharing framework. The Discussion Paper proposes four transparency and accountability mechanisms: Data Sharing Agreements; Public registers of Accredited Users and Accredited Data Service Providers; Annual reports on the data sharing system; Data breach scheme.

The University of Melbourne supports the proposal to require Data Sharing Agreements for all data sharing, and the proposal to make these public through a register maintained by the National Data Commissioner. We note, however, that some of the proposed mandatory terms of the Data Sharing Agreements may need to be further refined. For example, the Discussion Paper suggests that Agreements will include "a detailed description [that] will describe the data shared under the agreement".[7] Stated in these terms, the requirement is vague and too open to interpretation. To adequately support transparency, a detailed description should include a data dictionary listing every field included in the agreement, as well as information about the scale of the release. Clarity concerning the required content of Data Sharing Agreements is important for both ensuring transparency and for avoiding confusion that adds to the administrative costs borne by researchers and data custodians.

We also support each of the remaining three transparency mechanisms.

In addition to the identified mechanisms, the role of the National Data Commissioner as an independent oversight body will be integral to ensuring transparency and accountability in the data sharing system. The University of Melbourne urges that this role be appropriately resourced to ensure capacity for the Commissioner to contribute meaningfully to the needed awareness and cultural change that will underpin the maturation of an ecosystem that is highly motivated to get it right first time. As one small example, publicising positive stories about the value of data sharing will help maintain the social license for the use of this data.

Finally, consideration should be given to making it mandatory for decisions to *not* release or share data to be made public. This would encourage widened access to data for specified users, as well promoting a better understanding of the basis upon which decisions around data sharing are made.

12. Have we achieved the right balance between complaints, redress options and review rights?

13. Have we got our approach to enforcement and penalties right for when things go wrong? Will it deter non-compliance while encouraging greater data sharing?

Public trust in data sharing is buttressed by the protection given to individual rights. The Discussion Paper rightly notes that beyond the Data Sharing and Release legislation, measures to protect individuals are provided for in privacy, intellectual property, and consumer and competition laws[8], and that the legislation will allow for "internal and external merits review for accreditation decisions".[9]

Both the effectiveness of the legislative framework and public confidence in it will depend upon consistency across the various components of this framework i.e. Australian Consumer Law, Consumer Data Right and the Data Sharing and Release legislation. The ACCC's Digital Platforms Inquiry is reviewing the extent to which the Australian Consumer Law is 'fit for purpose' in allowing

[7] p.36.
[8] p.27.
[9] p.51

for adequate response to the misuse of consumer data. The protections for individuals provided in the Data Sharing and Release legislation should be developed in a way that is consistent with that complements that new measures proposed by the ACCC.

The University of Melbourne also emphasises the importance of adequate arrangements for redress, in view of the serious consequences that may be entailed in data breaches. We acknowledge the intention to establish a "comprehensive complaints mechanism" through the legislation.[10] It is important that the design of this mechanism is made clear, so that there is confidence it is sufficiently robust to maintain support for the data sharing provisions. We also note that complaint and redress provisions are only part of the overall protection mechanism. Privacy is not replaceable or repairable, if lost, so persistent education of data providers in international best practice for data management, sharing and release will be essential.

14. What types of guidance and ongoing support from the National Data Commissioner will provide assurance and enable safe sharing of data?

The National Data Commissioner should be empowered to evaluate and challenge operation of the data sharing and release scheme to ensure it continues to protect and promote public confidence in appropriate uses of data. This goes beyond providing advice on legislative proposals to include being able to challenge the operation of legislation in force.

The provision of technically sound guidance from subject matter experts whose primary expertise is in technical measures for privacy protection should also be a priority. The lack of expert technical guidance distorts public discussion around privacy and data sharing, and therefore has the potential to erode trust in the data sharing framework.

[10] Discussion Paper, p.45

Contributors to this submission*

Dr Sean Byars, Melbourne School of Population and Global Health

Ms Ximena Camacho, Director of Population Health Data Analytics, Melbourne School of Population and Global Health

Dr Chris Culnane, School of Computing and Information Systems

Associate Professor Graeme Hart, Surgery, Austin Hospital

Professor Anne Kavanagh, Chair in Disability and Health, Melbourne School of Population and Global Health

Professor Janet McCalman, Redmond Barry Distinguished Professor, Melbourne School of Population and Global Health

Professor Jeannie Paterson, Melbourne Law School

Professor Abigail Payne, Director, Melbourne Institute of Applied Economic and Social Research

Dr Megan Prictor, Research Fellow, Melbourne Law School

Associate Professor Ben Rubinstein, School of Computing and Information Systems

Professor Liz Sonenberg, Pro Vice-Chancellor, Digital & Data and Research Infrastructure & Systems

Associate Professor Mark Taylor, Melbourne Law School

*Note that while the researchers listed provided expert comment that informed the content of this submission, the submission ultimately represents the views of the University and not necessarily the views of each of these contributors.