Data Sharing and Release October 15, 2019

Dr Chris Culnane,
Contact Email: [redacted]

A/Prof Benjamin Rubinstein,
[redacted]

A/Prof Vanessa Teague,
[redacted]

This submission was prepared by the authors in a personal capacity. The opinions expressed in this submission are their own and do not necessarily reflect the official position of The University of Melbourne.

1 Data Sharing and Release

Prior to answering the specific questions raised, we examine in this first section broader concepts and assertions made in the discussion paper. Some statements are made with little or no justification, yet form the basis on which far reaching decisions around privacy, consent, and purpose have been made. We believe the techniques for privacy protection championed in the discussion paper are wholly inadequate. The discussion paper makes explicit reference to beneficiaries of data sharing, such as government, industry and the research sector, however the benefits to the public frequently claimed in the discussion paper are never cited with evidence. We believe that in special cases data collection and analysis can deliver significant societal benefit, however this does not warrant sharing-by-default. The discussion paper takes a one-sided view on consent, privacy safeguards, and purpose, with the resulting proposal unfit for purpose.

1.1 Defining Sharing and Release

The paper frames the public discussion of sharing and release in simplistic terms.

> "Until now, the public conversation in Australia has mainly been a binary one of open or closed data. Government agencies either kept data in-house or made it publicly available through data.gov.au or other websites."
>
> *Data Sharing and Release Legislative Reforms Discussion Paper [p.3]*

Such straw man framing is factually incorrect. The controversy over sharing of My Health Record Data (in particular, sharing with commercial entities) initiated one of the significant changes to MHR rules: the opportunity to opt out of secondary uses even while retaining a record. For other kinds of government-held data, significant opacity *about government sharing* means that Australians almost certainly underestimate the extent of sharing today.

Firstly, government departments are often *already* permitted to share data with other government departments. There are restrictions in certain circumstances, for example, the Australian census and sensitive medical data. However for most administrative data, extensive sharing is already taking place, as evidenced by the detailed list of data matching protocols for accessing multiple sources of administrative data as used by and published by the Australian Tax Office[1].

Second, data sharing regularly takes place between government and research organisations today, including with commercial for-profit entities. Our previous work demonstrating the re-identification of the 2016 public release of the MBS/PBS 10% dataset [1], showed the risks of ill-conceived data sharing and release policies. However, something not apparently known widely is that a similar linked dataset has been shared with research organisations [2], [3], [4].

Additionally, the PBS data has been shared with commercial organisations (*Hi Connections*) [5]. Apparently details of the sharing arrangement have been removed from the website, with just an email address provided as of writing. However, if we examine the `archive.org` [6] record of [5], one can still find a more detailed description for the sharing taking place with *Hi Connections*:

> "Hi Connections has provided services to the pharmaceutical industry since October 2004. This includes longitudinal patient tracking back to June 2002 to identify therapeutic pathways pre and post initiation; persistence and compliance; co-medication, dosing and titration; time to event and authority indication analyses. Analyses are based on a 10% PBS patient sample pioneered by Hi Connections and supplied by the Department of Human Services (formerly Medicare Australia)."
>
> *Web Archive 19th April 2015 of [5] available from [6]*

Some time in mid-2016 details of the data sharing that had taken place with *Hi Connections* were removed and replaced by just an email link. This unequivocally demonstrates that the sharing of what most would consider sensitive data has taken place with commercial organisations previously.

We can assume that the data is still being made available to *Hi Connections* given that it is still being referenced in Academic publications [7]. However, the company has no website and there is no public statement about how the data is being held, analysed, or shared. It is also clear that the current framework is a long way from a binary closed or open setup.

A far clearer definition of use, sharing, and release is required, based primarily on four different categories:

Use Government use of data it has collected. Such usage must still be bound by strict purpose constraints; usage outside of the expected or stated purpose would still require some form of consent.

Access Allowing researchers to perform approved queries on data in controlled lab conditions.

Share Sharing of the data outside the organisation but bound by a legal agreement that both limits the usage, but also maintains control and guardianship of the data with the original receiving entity. This would limit the type of sharing that takes place with trusted research partners for example, who would be bound by a time-limited contract to perform a particular type of analysis. In these instances consent would be required in most cases as it would constitute a secondary use.

Release The public release of data for uncontrolled usage, for example `https://data.gov.au`. The type of data that is suitable for such release is extremely narrow, with detailed unit-record level data about individuals never suitable for release.

[1] `https://www.ato.gov.au/General/Gen/Data-matching-protocols/`

1.2 Primary vs. Secondary Uses

It is remarkable that primary and secondary uses of data are not even mentioned in the discussion paper. APP 6 — as defined in the *Privacy Act 1988* — covers use and disclosure, based on the distinction between primary and secondary uses.

> "6.1 If an APP entity holds personal information about an individual that was collected for a particular purpose (the primary purpose), the entity must not use or disclose the information for another purpose (the secondary purpose) unless:
>
> (a) the individual has consented to the use or disclosure of the information; or
>
> (b) subclause 6.2 or 6.3 applies in relation to the use or disclosure of the information."
>
> *Privacy Act 1988*

The distinction between primary and secondary use is closely related to the paper's discussion of consent. The approach proposed in the discussion paper is the worst of both worlds: the distinction between uses has been lost and the requirement of consent significantly eroded.

One benefit of distinguishing between primary and secondary uses of data is to shift the consent boundary from the point of access to the point of secondary use. This ensures maximum uptake of essential services, which is both in the interests of the individual and a productive, inclusive and safe society.

By removing the distinction between primary and secondary uses, the consent boundary shifts to the point of delivery or access. An individual still holds consent powers, except in this case, if they choose not to consent they have to forgo access to a service they are entitled to, or risk penalties for non-compliance with mandated data collection. For example, when a commonwealth-funded work-for-the-dole provider was criticised for uploading clients' details to a public facebook page,[2] it replied that "We do not believe that this is a breach of confidentiality ... participants give ... media consent when they commence as a participant." So it is not possible to access the service without giving consent for excessive sharing of information.

This pattern creates a risk that members of the public will not make use of essential services or will respond with misinformation because they fear how their data will be used. Those services are provided by society on the basis that engagement with them, by those who need it, is collectively better than if they did not make use of such services. For example, if a person with a stigmatised infectious disease chooses not to seek medical care for fear that their condition will be exposed, then the consequences impact not only that individual, but also the rest of their community through preventable infection of others.

1.3 Consent

The discussion paper's treatment of consent is completely inadequate. Section 4.6 in particular asserts that

> "Requiring consent for all data sharing will lead to biased data that delivers the wrong outcomes."
>
> *Data Sharing and Release Legislative Reforms Discussion Paper [p.33]*

[2] https://www.abc.net.au/news/2019-04-26/nt-centrelink-client-details-published-on-public-facebook-page/11049804

and even more alarming,

> “If we required consent, then data would only be shared where consent was given. This will skew the data which is shared, leaving it unfit for many important purposes in the public benefit; it also runs the risk of leading to flawed policy and research which impacts negatively on society.”
>
> *Data Sharing and Release Legislative Reforms Discussion Paper [p.33]*

These arguments present an equivalence between overriding consent and getting unbiased statistics, which is false for the reasons given above—people may withdraw their data by withdrawing from, or misleading, the service provider, if they do not think their privacy will be adequately protected.

Furthermore, the arguments above present a false dichotomy between *requiring* consent and *not getting enough* consent for useful scientific studies.

Existing data collection practices and ethics policies across the academia, medical and research sectors require the provision of consent for the collection of data. Today, every survey conducted by an academic requires participants to consent and be given the ability to freely withdraw or drop out at any time. If the above discussion paper equivalences were factual, all of those data collection activities, and the multitude of research papers based on them, would be too biased to have value. Similarly of much of international science. That is not the case—competent researchers understand that consent may result in participants refusing to provide data and that has to be factored in to their experimental designs, survey sampling strategies, and statistical analysis in determination of the significance of results.

On public support (or lack thereof) for consent, the discussion paper is inconsistent. Any relaxation of requiring consent contradicts the claim of only releasing data if the Australian community supports it:

> “The Data Sharing and Release legislation will not allow public sector data to be shared if it is considered too sensitive, for example, because it would threaten Australia’s national security or because the Australian community does not support it.”
>
> *Data Sharing and Release Legislative Reforms Discussion Paper [p.14]*

Individual consent and social license are conflated by the discussion paper in order to justify invasion of privacy on the grounds of projections of societal benefits. The best way of determining whether the Australian community supports a particular application of data sharing is to ask for their consent, not for an unaccountable individual to make that decision for them.

The purpose of asking for consent is both a test of support and a requirement by the requester to justify why they need access to the data and to be able to justify why the benefits outweigh the risks. The argument for access to higher utility data, for example, identifiable data, is made during the consent process. If a researcher cannot make a compelling enough argument to obtain consent then they need to either reconsider their analysis, or retry their articulation of societal benefit.

Irrespective, the consent point will still exist: there will still be the possibility of disengagement from a service to deny consent for data sharing. The pursuit of more data could very well lead to serious unintended consequences, of both less data and less usage of essential services. Total disengagement is far worse than denial of consent; at least with denial of consent the size and nature of the denial can be modelled so that some results can still be obtained and a confidence can be calculated on the outcome.

1.4 Compatibility with Human Ethics

Removing the requirement for consent appears to create a conflict with ethical requirements of research as defined in the *National Statement on Ethical Conduct in Human Research 2007 (updated 2018)* which reflects societal expectations and international norms for research. Principle 2.2.1 states:

> "The guiding principle for researchers is that a person's decision to participate in research is to be voluntary, and based on sufficient information and adequate understanding of both the proposed research and the implications of participation in it."
>
> *National Statement on Ethical Conduct in Human Research 2007 (updated 2018) [p.16]*

The proposed data sharing framework directly contradicts the guiding principle of voluntary participation. Whilst there are exceptions to requiring consent, for example, in the case of limited disclosure, the requirements to obtain such exception would not be met in the context of this data sharing, for example,

> "2.3.1 a) there are no suitable alternatives involving fuller disclosure by which the aims of the research can be achieved"
>
> *National Statement on Ethical Conduct in Human Research 2007 (updated 2018) [p.20]*

There clearly are alternatives, e.g. seeking consent. The discussion paper argues not that asking consent in impossible, it argues that asking for consent might result in people saying no, which is not a legitimate reason to not ask for consent. The closest exception is covered by an opt-out approach which can be deemed legitimate if a number of requirements are met, including:

> "2.3.6 c) the research activity is likely to be compromised if the participation rate is not near complete, and the requirement for explicit consent would compromise the necessary level of participation"
>
> *National Statement on Ethical Conduct in Human Research 2007 (updated 2018) [p.21]*

However, it should be noted that such opt-out approaches still provide a pseudo-consent mechanism, since reasonable attempts must be made to notify the participant in advance and for them to be given the option to opt-out and not have their data included.

Secondary usage of data, even public data, again presumes the requirement for consent:

> "3.1.52 Unless a waiver of the requirement for consent is obtained, any research access to or use of publicly available data or information must be in accordance with the consent obtained from the person to whom the data or information relates."
>
> *National Statement on Ethical Conduct in Human Research 2007 (updated 2018) [p.37]*

It is unfortunate that the discussion paper does not reference the *National Statement on Ethical Conduct in Human Research*. While a major motivation for the proposed legislation is to provide researchers with more data, most genuine scientific researchers will not be able to use such data as currently proposed. The National Statement is the primary document laying out out the principles under which researchers can access data.

Should the legislation be introduced and passed it will mark an abandonment of one of the fundamental principles of research — voluntary consent — and will substantially weaken protections of Australians' private data at a time when most other democracies are introducing stronger protections.

2 De-identification and re-identification

The discussion paper shows an astonishing faith in the likely success of de-identification, despite noting in passing that many people (including us) tried to explain that de-identification of detailed unit-record level data does not work. A detailed record about an individual, even if it has had the most obvious identifiers (such as date of birth and place of residence) removed, is still likely to be easily re-identifiable.

Some of the most controversial data-sharing questions concern the sharing without consent of information about a person who is *not explicitly identified* but nevertheless *easily identifiable*. The Office of The Australian Information Commissioner found in their investigation into the MBS-PBS 10% sample data breach that patients were not 'reasonably identifiable' under the Privacy Act, despite being easily and confidently identifiable from a few pieces of information.[3] We do not agree that easily-identifiable detailed personal information should be excluded from the protections of the Privacy Act. We do not attempt to argue here the legal question of whether it already is protected: one interpretation is that the OAIC were simply wrong, and that easily and confidently identifiable personal information is 'reasonably identifiable' and hence protected by law even if the identity of the person is not explicit. Another interpretation is that it is not protected now, but should be, for example by amending our Privacy Act to protect data about an 'identifiable' rather than only a 'reasonably identifiable' person. This question will continue to be central to both government and commercial data sharing.

It is important to note that the risk of re-identification is not only a problem for open data, though the discussion paper mentions it only in that context. Many commercial entities might have a strong incentive to re-identify data (for example, health or financial data) if it has been given to them in 'de-identified' form but individuals were actually easily identifiable. Privately-held commercial datasets (such as as a person's banking and health insurance information) would greatly aid in re-identification of related government datasets. The result could be significant harm to the individual whose data had been shared. It might include exclusion from credit, insurance or employment.

3 The Data Sharing Principles

The Five Data Sharing Principles are little more than a renamed version of the Five Safes Framework. In our previous submission we criticised the Five Safes Framework for lacking rigour and failing to provide effective privacy protection, and recommended at the very least switching to a five risks framework. The whole purpose in switching to a risk based approach is that the discussion becomes framed about what might go wrong, not what might go right. While it may appear that the Data Sharing Principles are now a risk-based framework, the wording in both the discussion paper and the *Best Practice Guide to Applying the Data Sharing Principles* is evidence of a superficial rebranding without meaningful change in focus. If we examine the breakdown of the five principles in the discussion paper they remain focused on safety not risk:

[3] https://www.oaic.gov.au/privacy/privacy-decisions/investigation-reports/mbspbs-data-publication/

Project	Data sharing is for an appropriate project or program of work
People	Data is only available to authorised users
Setting	The environment in which the data is shared minimises the risk of unauthorised use or disclosure
Data	Appropriate protections are applied to the data
Output	Outputs are appropriate for further sharing or release

Data Sharing and Release Legislative Reforms Discussion Paper [p.30]

None of the above indicate an evaluation of risk and none even superficial box-ticking for risk management. Each principle projects the existence of low risk outcomes by default, which is not the same as judging risk with an open mind of either approval or rejection of sharing or release. If the principles were genuinely risk based they would have been worded as such, for example:

Project	Risk the project or program of work deviates from that planned or is insufficiently described to have been effectively evaluated
People	Risk of unauthorised access, or malicious usage by an authorised individual
Setting	Risk of compromise of the environment in which the data resides, for example, malware or hacking attack
Data	Risk data reveals information beyond what had been intended to be shared or released, for example, re-identification of an individual
Output	Risk of releases, even in aggregate form, lead to direct re-identification or assist in the re-identification of a third-party dataset, or unwanted attribute reconstruction

The difference between the two is crucial, the latter requires the process to evaluate a risk and then justify that evaluation, and if necessary, implement mitigating strategies. It follows the same process of cyber security risk evaluation in that you assume the existence of an attacker (with a specific threat model for the attacker and trust model for cooperating parties) and evaluate the risk of an attack succeeding, as opposed to the proposed principles focus on predicting a safe outcome.

3.1 Best Practice Guide to Applying the Data Sharing Principles

The discussion paper refers to the *Best Practice Guide to Applying the Data Sharing Principles*, released earlier this year. It appears that in rewording of “safe” to “principle”, contradictions were introduced:

> “While each Principle can be considered independently, all five Principles need to be considered jointly to evaluate whether a particular instance of data sharing is a safe arrangement”
>
> *Best Practice Guide to Applying Data Sharing Principles [p.12]*

In replacing the word “Principles” with “Safes”, the statement would become consistent with the Five Safes Framework. The present wording is inconsistent: a risk-based approach cannot by definition result in a safe arrangement. Such an approach can only lead to lower risk arrangements.

Another example statement in the guide:

> "Controls should be based on a realistic assessment of the likelihood and consequence of a risk occurring and be made in the context of organisational risk tolerance, rather than based on hypothetical worst case scenarios."
>
> *Best Practice Guide to Applying Data Sharing Principles [p.12]*

This would be appropriate if the data being shared was about the organisation itself, since it is the organisation that will incur the cost of any loss. However, when sharing other peoples' data, the risk tolerance of the organisation is a distant secondary concern. Without a privacy tort in Australia, an organisation, particularly a government agency that can mandate data collection, can have an extremely high tolerance of risk of loss of other peoples' data because the consequence for it is close to zero. No significant consequences have been faced by the government after releasing identifiable MBS/PBS records of 2.9 million people online. As far as we know, the affected people have not even been notified. Nor has Public Transport Victoria suffered significant sanction despite the Office of the Victorian Information Commissioner's finding that their sharing online of 3 years of Myki touch on and off events breached Victoria's Privacy Act.[4] The privacy cost is borne by the public — the affected data subjects — and the data curator suffers no loss. In such an environment the risk tolerance for the government entity may be considered almost unbounded.

The guide also fails to address the problem of interdependence within the different principles/safes. Almost all of the principles are co-dependent on people. For example, evaluating whether the project is in the public interest is only valid if the project proposal, which is written by the intended data recipient, is an honest reflection of what they intend to do. If that person lies and claims they want to complete a project in the public interest, but in fact are going to do something for personal financial gain, then the integrity of the project phase is compromised. As such, it is not possible to have a lower level of control over people than project - as happens with aggregated data, since any protection cannot exceed that offered by the control mechanisms applied to people. Such dishonest self assessment occurred in the case of Cambridge Analytica. The problem is exacerbated when technical protections around the data are reduced, resulting in far greater dependence on people. Notably the Data Sharing Principles do not stipulate any particular technical protections even though broadly accepted safeguards exist for commonly adopted threat models.

The description of the data principle is not fit for purpose: It does not even mention the only accepted rigorous privacy model, differential privacy, which could be applied in the aggregate setting for which high data control is required. Differential privacy can hardly be described as new: the U.S. Census Bureau has committed to releasing 2020 census statistics in a differentially private form [8]. Yet it isn't mentioned in the Australian document describing best practices in applying data sharing principles. That omission seriously calls into question the validity of the entire guide.

Of even more concern is that it refers to the *De-Identification Decision Making Framework*. This framework recommends inadequate methods of protection which, if applied trustingly to longitudinal data, are likely to result in releases of easily identifiable information. Not only is its discussion of longitudinal data entirely inadequate, but it gets the definition of k-anonymity wrong, leading to an incorrect application of its principle and incorrect calculations of disclosure risk. We recommend that the guide be completely reworked.

[4] https://ovic.vic.gov.au/wp-content/uploads/2019/08/Report-of-investigation_disclosure-of-myki-travel-information.pdf

4 Questions

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

The current definitions conflate sharing internally with sharing externally, which is compounded by not distinguishing between primary and secondary uses. If the discussion paper had been consistent with the *Privacy Act 1988*, and distinguished between primary and secondary uses, then the problem would not be so severe. Primary uses would be covered under a usage provision, and all secondary uses, whether they be within the entity or external, would require passing a consent hurdle.

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

The methodologies for release are often flawed with state and federal government approaches not consistent with best practice. The release of the MBS/PBS dataset openly [1], as well as the intended release of the Victorian Myki dataset via open data channels [9], demonstrate a disregard for privacy protection. It is quite simple: unit record level data about individual members of the public should never be made openly available [10]. Aggregate data should be protected via rigorous mathematical privacy technologies, for example differential privacy as adopted by the U.S. Census Bureau [8] among many in government and industry internationally. Data that refers to elected representatives, governments actions, and the environment should be made openly available.

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

The framework will achieve more streamlined sharing processes with greater quantities of data being shared in the short term. However there will be a significant reduction in privacy for the Australian public. The proposed Data Sharing Principles are the Five Safes Framework under another name and are no better than the original framework which is flawed. Privacy protecting releases of data is not a solved problem in general, and it is not going to be solved by streamlining the process to enable greater sharing. Over-sharing now will make protecting future data releases more difficult and potentially erode social license for legitimate sharing, potentially limiting data sharing in the long term.

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

The name reflects the substance of the discussion paper. However significant concerns about the discussion paper remain.

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

We support precluding compliance and national security uses. However, the purpose test is only required due to the removal of consent. If consent were present, then the public would be involved in the purpose test instead of only the data custodian and data user. Assigning uses of data into three categories is too coarse: for example some R&D purposes represent legitimate research in the public interest, while others have public interest as cursory secondary motivation. There is no threshold for public interest present in the purpose test.

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?

As previously discussed the Data Sharing Principles are not appropriate, and the existing legislative protections are not sufficient. There is no legitimate reason for unit-record level public data to be shared with commercial entities without the consent of the data subject. Additionally, research uses of data should be required to publish papers in open access locations with any outcomes (patents, tools, designs) being placed into the public domain for the collective benefit of all.

One of the overarching problems with the discussion is the conflation of profit and productivity. This conflation pervades almost all the reports produced around open data and data sharing. Productivity is being used as cover for profit. Productivity is a measure of efficiency of output, it is considered to be beneficial to society because at an abstract level an increase in productivity permits either an increase in earnings, or an increase in available time. By contrast, an increase in profit may only benefit the owners of the company. As such, the collective interest in sharing data for the purposes of generating greater profits for the few is not automatically justified by default.

The type of data that more clearly leads to productivity increases is utilitarian data. For example, data that allows someone to move between locations A and B quicker, or to find service C in less time. Much of this type of data is not about individuals. The current location of every Tram, what the routes are, the GPS co-ordinates of the stops, the time table, etc. are all examples of utilitarian data that could increase productivity and drive innovation. The travel patterns of every Myki card in Melbourne does not lead to a productivity increase — it could lead to policy improvements that could be achieved without sharing data outside of the government itself.

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

The Data Sharing Principles are not a risk based framework, it is the Five Safes Framework with a superficial title change. They are still resolutely focused on the notion of safe release, which is incompatible with a risk framework. The Data Sharing Principles is a governance framework that does not stipulate any particular trust models. As such it is easy for a data custodian and data user to satisfy the framework by self-reporting sharing as desired. This does not meet public expectations and the framework should be completely rewritten with international, peer-reviewed, best practice in mind.

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

The guide could be improved by being rewritten to be a technically sound risk-based framework. There are well-known examples of similar risk frameworks available, for example, the OWASP Risk Rating Methodology [11].

9. Do the safeguards address key privacy risks?

The safeguards do not address privacy concerns, they are interdependent and suffer from the same shortcomings as the Five Safes. There remains over reliance on evaluating people, a task which no one knows how to do efficiently or effectively. The safeguards do not even mention current best practice — differential privacy for release — but do reference poorly defined guides using techniques that are unfit for purpose, such as the *De-identification Decision Making Framework*.

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

The level of coverage offered by the core principles is heavily dependent on the substance of them, which is not currently defined. For example, "skills and capabilities to protect, manage and use data" is a laudable goal, but how is that going to be delivered, particularly in an environment that so far has not delivered such capabilities?

The second principle is particularly troubling since it is applicable only to those handling personal information. As was demonstrated by the Victoria Department of Transport's response to the Office of the Victorian Information Commissioner's report [12] into the Myki breach, there is far from uniform agreement of what constitutes personal information. Rather than being determined by a legalistic conceptualisation of personal information, the principle should apply to anyone handling data about, or derived from, individuals. The form of the data, whether it is unit-record level or aggregate should not matter, since the risk to individual privacy remains.

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?

The details are not sufficient to be able to make such a judgement. For example, "A detailed description will describe the data shared under the agreement" could mean many different things. It should mean a data dictionary listing every field included in the agreement, as well as information about the scale of the release. However, it is easy to see how this could be interpreted as just a one paragraph description of the nature of the data.

Furthermore, agreements should be made public and open to challenge prior to any transfer of data. Transparency without accountability has little value, and there is no benefit in waiting for a failure to occur before reacting. There is little to no detail of actual accountability measures. What will it take for an entity to lose accreditation? Given that breaches are likely to be internal only, i.e. unreported scope creep in a reported project to run additional analysis, what protection will be provided to whistle-blowers? What right to public appeal are the accreditation's or agreements? What financial and privacy redress measures are available to an individual whose data is breached?

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

There appears to be no redress for an individual whose data is compromised in a breach. In fact, the legislation appears to largely ignore the individuals to whom most of the data will relate. They no longer are asked for consent, they don't appear to have any rights under the accreditation scheme to oppose accreditation or sharing agreements, and they have no redress when things go wrong.

13 Have we got our approach to enforcement and penalties right for when things go wrong?

The current approach is too soft, and fails to create an ecosystem that is highly motivated to get it right the first time. Privacy is not something that can be recovered. It is not replaceable or repairable—the loss will be felt for an extended period of time, potentially for the rest of the person's life. As such, it is not appropriate to have a redress scheme that seems to be based on trial and error. Such an approach encourages entities to have little regard in their initial approach to protecting and safeguarding data. The enforcement approach will only encourage investing in suitable protections following a breach, and even then, it could take multiple breaches before there is any meaningful motivation for protecting privacy.

If the intention is to incentivise good behaviour the penalties have to kick in immediately, and they have to be severe enough to make it more desirable to comply than to just take the penalty.

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

The first priority should be technically sound guidance from subject matter experts whose primary expertise is in technical measures for privacy protection. The current guidance is not fit for purpose. This is distorting the wider discussion about privacy and data sharing and risks creating an existential threat to data sharing in all forms. Until there is acceptance that data is a digital representation of one's physical self, which should be afforded the same rights and protections as that person, the conversation will not progress in a constructive fashion.

References

[1] Chris Culnane, Benjamin I. P. Rubinstein, and Vanessa Teague. Health data in an open world. *CoRR*, abs/1712.05627, 2017.

[2] Australian Longitudinal Study on Women's Health (ALSWH). ALSWH and MBS/PBS Linked Data Notes. http://www.alswh.org.au/images/content/pdf/Extra_Files_for_Website/Access_Data/Doc%20I%20MBS%20PBS%20request%20form%20updated%20160215.pdf. Accessed: 2019-10-12.

[3] Australian Bureau of Statistics (ABS). 1700.0 - Microdata: Multi-Agency Data Integration Project, Australia. https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/1700.0main+features10002Australia. Accessed: 2019-10-12.

[4] Sax Institute. Technical Note: Linked MBS and PBS data. https://www.saxinstitute.org.au/wp-content/uploads/Technical-Note-Linked-MBS-and-PBS-January-2017.pdf. Accessed: 2019-10-12.

[5] The Department of Health Australian Government. Sources of data for use in generating utilisation estimates. http://www.pbs.gov.au/info/industry/useful-resources/sources. Accessed: 2019-10-12.

[6] archive.org. Archive of http://www.pbs.gov.au/info/industry/useful-resources/sources for 19th April 2015. https://web.archive.org/web/20150419114027/www.pbs.gov.au/info/industry/useful-resources/sources. Accessed: 2019-10-12.

[7] Sarah Louise Sheridan. Assessing vaccine effectiveness of publicly funded vaccination programs in queensland. 2017.

[8] John M Abowd. The us census bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2867–2867. ACM, 2018.

[9] Chris Culnane, Benjamin I. P. Rubinstein, and Vanessa Teague. Stop the open data bus, we want to get off. *CoRR*, abs/1908.05004, 2019.

[10] Office of the Victorian Information Commissioner. Protecting unit-record level personal information. https://ovic.vic.gov.au/resource/protecting-unit-record-level-personal-information/. Accessed: 2019-10-12.

[11] Open Web Application Security Project. OWASP Risk Rating Methodology. https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology. Accessed: 2019-10-12.

[12] Office of the Victorian Information Commissioner. *Disclosure of myki travel information - Investigation under section 8C(2)(e) of the Privacy and Data Protection Act 2014(Vic)*. 2019.