# Submission in response to the Data Sharing and Release Legislative Reforms Discussion Paper, September 2019

**Professor Kimberlee Weatherall, The University of Sydney Law School**
**Libby Young, Researcher, The University of Sydney Law School**

We welcome the opportunity to comment on the latest Discussion Paper ('the Paper') regarding the Data Sharing and Release Legislative Reforms. We hope the early phases of this consultative process - explaining, listening and iterating with different stakeholders - represent the beginnings of an open, collaborative and ethical culture around how data is used for public benefit in Australia. The long-term success of these reforms depends on continued hard work to build such a culture, and ongoing iteration as we learn more about the operation and impact of these reforms, and sharing and use of government-held data.

With all this in mind, we focus on three areas in this submission: 1) the conceptual risk management approach and model that informs the Paper; 2) the two ongoing debates around consent and commercial use; 3) how the theory of these reforms can be translated into practice.

## 1. The conceptual model

We support the twin overarching goals guiding the Paper and this process of reform:

- to share publicly held data more effectively, **to benefit the Australian public**, and
- to do this in ways which actively **foster trust from the Australian public**.

At the same time, we note that the Paper has, at its core, a model that builds on the the data sharing risk management model, the 'Five Safes'. The Paper suggests that the trust of Australians in data sharing depends on maintaining privacy and security, and keeping sharing (beyond government in particular) within the bound of community expectations. We certainly agree that safeguards, security and community expectations are important (we agree, for example, that sharing for commercial purposes, or for compliance/enforcement, should not be enabled by the legislation).

But we advocate for a more positive vision, and a more holistic understanding of the context of data sharing. The Paper's current focus on safety and efficiency, while important in today's climate of mistrust exacerbated by high profile cases of data misuse, also risks entrenching a normative culture around data sharing in Australia that is paternal and pessimistic. This risks limiting *at the outset* the potential of these reforms and data sharing in general to enrich the country's imagination and vision, using new tools of data and decision-making for the benefit of

diverse groups with diverse needs. We need to remind ourselves that with new tools we can imagine new possibilities.

In thinking about what data sharing can achieve, Australia **can and should promulgate a more positive and creative vision, aspiring to deliver positively beneficial outcomes for the people of Australia through data sharing,** not merely smoother service delivery or 'better targeted' policy (a phrase which could easily be read as suggesting the best use of data is to restrict the availability of/access to government services). We should seek to enable a public service **culture of improvement** beyond efficiency or 'better targeting' alone**.** Efficiency is one, but not the only or best way to improve a policy or service. Indeed, the pursuit of efficiency on its own can *degrade* the quality of services for citizens, as the example of Robodebt has shown.[1] We would hope that governments would actively seek (creative) ways that data sharing could, for example, improve the lives of people resident in Australia; or improve society, or the environment.

We should also seek to support the **agency and autonomy** of Australian citizens and residents**.** This means involving the public, or affected sections of the public, in decision-making around the sharing or use of data. Part of the longer-term project here can be to undertake data sharing in ways that help build the public's capacity to engage with the challenging issues around data use, rather than exacerbate the impression that extensive data collection, sharing and use is beyond the public's control.[2] We suggest, for example, that those involved in testing the reforms seek out potential local projects where communities or groups could be actively involved in deciding what data is needed to solve a local problem, and how it can be shared and used.

We would also contend that safety and security of data are necessary, but not sufficient conditions for building public trust in data use. Rather than focusing only on staying within the bounds of what we assume people will accept, can we seek to foster trust by understanding and addressing contexts for data sharing where individuals perceive there is no real choice not to share, or the choice not to share is unclear, or it is too costly not to share?

Another comment should be made about the Paper, and that is the absence of a discussion about data *use*, including current developments in data use around that set of technologies commonly known as artificial intelligence/machine learning/automated and intelligent systems ('AI' here for simplicity's sake). We appreciate that a decision may have been made not to discuss questions of data use, as being, perhaps, a 'next stage' and a responsibility of a data recipient rather than the data sharer. We are also aware of other ongoing processes relating to data use and AI, such as current efforts to develop a Commonwealth AI Ethics Framework.

---

[1] We do not dwell further on Robodebt here, noting that this Paper expressly excludes data sharing for the purpose of compliance and enforcement. Nevertheless, Robodebt has resonance for the broader culture of data sharing across government agencies.

[2] See comments below regarding consent and alternative mechanisms for community approval of data sharing and use.

However, we think not discussing data *use* at this point in the consultation creates a weakness: much of the analysis is premised on an assumption that data sharing will occur for limited research or policy projects that end with some report. It does not discuss the fact that in a context where AI is used, data sharing can lead not only to *outputs* (reports, analyses etc) but to *outcomes* (real world impacts and decisions that affect the rights or interests of individuals). For example, AI tools may be used to make or implement policy decisions. It is not clear that the 'Five Safes' model (or the version elaborated in the Paper) adequately addresses the risks that arise when outcomes, rather than outputs, are a potential consequence of data sharing.[3]

Nor is it sufficient to say that the appropriateness of outcomes is a matter for a data recipient rather than an entity that is sharing data. A data custodian is more likely (albeit not guaranteed) than a recipient or requester of data to have a fine-grained understanding of the strengths and weaknesses of the data, how it was collected, and what can and *cannot* be inferred from the data. Thus some responsibility for assessing potential uses of the data, including outcomes, should be considered by the data custodian when deciding whether or how to share. If for no other reason than that that information needs to be properly conveyed with data transfer.

## 2. The Gordian knots: consent and commercial use

We now focus on two complex and high stakes questions rightly highlighted by the Paper and discussion workshops; those of **consent**, and **commercial purposes.**

**CONSENT**

This latest phase of the consultation includes the suggestion that making consent obligatory for all data sharing may contribute to biased or incomplete data sets, and that this may in turn limit the benefits of data sharing for citizens and residents, in particular vulnerable groups. (We note the Open Data Institute's idea of 'data wastelands'[4] as a potentially useful concept here). To address this issue, it has been suggested consent for data sharing is not always required.

We agree it is important to distribute the benefits of data-sharing as fairly as possible, and that we are currently too far away from achieving that goal. We commend the consultation's raising of this issue. But we strongly disagree that doing away with consent to address this issue is a

[3] Notably the Five Safes framework itself, perhaps as a result of being written prior to the current explosion in AI techniques and their use, does not address *outcomes*, but focuses on *projects* and *outputs*, reflecting quite a different model of the data/policy lifecycle than we see today. The Australian Computer Society in a recent White Paper suggested that in an AI context, additional considerations include **safe organisations** (referring to the systems, processes and governance employed by an organisation to ensure the Five Safes Framework is applied throughout the project and with the long-term management of data and outputs); **safe outcomes** (referring the ultimate uses of the project outputs); and **safe lifecycles** (to the time sensitivity of data or outputs).

[4] https://theodi.org/about-the-odi/our-vision-and-manifesto/our-theory-of-change/#1531394343108-b226e61c-833d

sustainable solution, a responsible approach, or in keeping with the values of a democratic society.

If vulnerable groups benefit less from data sharing because of issues with consent, then **the consideration of consent, its function, and how those functions could be addressed, should become more important, not less**. Any question of consent must keep front of mind its role in human relationships and political legitimacy - to establish and maintain trust, to embody and uphold democratic values, and to promote and protect the agency and autonomy of citizens and residents, both as individuals and as groups.

As a way forward on this vexed question, we make the following observations:

1. Consent supports individual agency and autonomy. But the process of seeking consent also provides both a measure of accountability (the entity requesting data must reveal their intended data use to the affected individual or group) and an opportunity for the data subject to decide whether the benefits of sharing data outweigh the costs (the assumption being that they are in a better position to know, perhaps, than the person requesting data). If consent is not possible, or will undermine the purposes of a project, we can still look for means of creating accountability (of the data recipient to affected individuals or groups) and mechanisms for ensuring that those making the decision to allow the data use take all the costs into account, as well as the benefits.

2. The choice is thus not a binary one, between individual consent and public interest override by the data custodian, with no other possibilities. **Rather than frame the issue as one of "consent vs override consent in the public interest", the question instead is: what is the appropriate mechanism for consent or its functions in each case?** There is a spectrum of other options for creating accountability and ensuring that data sharing is done with sensitivity towards the interests of people about whom data is being collected, used, or inferred. For example, in considering whether data should be shared to enable investigation of the needs and activities of a population subgroup, entities that represent that subgroup could be consulted. In the case of data collection that affects particular communities, the community can be consulted without requiring every individual in the community to give individual consent. In the case of population-level, sensitive/high-impact projects, other kinds of consultation are available, including, for example, citizen juries or other discussion exercises. Accountability can be created by requiring data recipients to publicly record their data use in a way that can be assessed by affected individuals or groups.

3. In this connection we note that interests in data are not purely individual, but also communal: data sharing affects groups within society, and may do so even where individual privacy is unaffected. This is further reason not to approach questions of consent as binary. In some cases, consulting with a community may be appropriate where a community is particularly affected by a project or proposed use. In this respect,

we note that communal interests in data are core to discussions around Indigenous data and Indigenous Data Sovereignty: and that some of these insights may well be useful in other contexts.

4. Additionally, if datasets are biased or incomplete because of issues of consent, any attempt to address this needs to understand *why* there is an issue. Is it because some individuals or groups have no access to technology? Because they have no time to consider consent? Because there is a lack of trust in the data user? Because there is a lack of trust in how the shared data will be used? In all these cases, overriding the need for consent risks further eroding trust and so lowering the quality of data more. In short, there is a serious risk that removing consent may make the issue the consultation rightly identifies worse not better.

5. Lastly, a model built around consent has proved unsatisfactory in the context of data collection, sharing and use within the private sector: endless privacy policies that stand as proxy for consent have not engendered significant trust in large social media companies, for example. Where provision of data is a condition of receiving a service, "consent by contract" is neither real, nor builds trust. Similarly there is little reason to think that consent will address concerns that might arise around government data sharing. We need to be sensitive to the fact that people in need of government assistance will have little ability to refuse consent for data use.

**COMMERCIAL PURPOSES**

We note the Paper's clear proposed framework for *when* data sharing is acceptable: for government policy and programs, for research and development, and service delivery. We also note commercial use is *excluded* as a fourth potential justification. Again, we agree that sharing publicly-held data for private gain will in many instances erode trust and reduce fairness, and commend the consultation for recognising this issue. We live in an environment of runaway data collection and use by private entities. There is a wariness of and in some cases outright fear of data being held by private entities, processed in opaque and uncontestable ways, and used for private gain without clear benefits beyond that.

But again as with consent, we question the suggested approach to address the issue here, and ask if there is a different model to be explored. **If we are to allow ourselves to use these new tools to imagine new possibilities, what if data sharing for commercial use can be *done better*, and support a fairer and safer society?** What if there are alternative models to '*data as a resource to be handed over with no obvious return to society*'? What if instead we were to embody the spirit of sharing more fully, and think about mutual obligations and benefits?

As a net importer of 'big tech', Australia needs to think differently on this question. And perhaps an early indication of how all this could shift has already surfaced in the principle of reciprocity[5]

[5] https://treasury.gov.au/sites/default/files/2019-03/Review-into-Open-Banking-_For-web-1.pdf

proposed in Australia's new Open Banking Framework. In this first expected implementation of Australia's new Consumer Data Right, big banks sensitive to conceding a competitive advantage to big tech have advocated for reciprocal sharing. **In the same spirit, what new possibilities are there for reciprocity between citizens and businesses here in Australia in ways that support a fairer and safer society?**

## 3. Putting theory into practice - how will this work?

Lastly, looking ahead to the next phases of this process, we ask **how will this all be applied**? We consider three areas in particular: testing; teaching; and values.

**TESTING**
As already emphasised, we commend the iterative approach of this consultation so far. To continue this beyond the theory and drafting, we hope the roadmap planned by the consultation team will include a prototyping plan, supporting implementation that is iterative, and an ethos of learning by doing. The Data Commissioner could perhaps play a valuable role in guiding and enabling this sort of approach.

Questions a prototyping plan may consider include:

1. With regards to the earliest phases of implementation, **where will testing start?** Given the much higher risks and impacts involved in sharing personal data (de-identified or not), the first tests for these reforms should start with data that is *not* about people. For instance, data about weather or infrastructure.

2. With regards to the first prototypes using data about people, **how will the communities affected by prototypes be involved in these processes?** As noted above, we would suggest local projects as one way to start. Not just so individuals and groups can consent or support 'user testing' for the purpose of data sharing to improve service delivery, but so these groups can also actively participate in prototyping processes for all types of data sharing proposed in these reforms, including for policy making and research. This should include taking part in the decisions about how personal data is used (**with the ability to say no**), and the opportunity to question, learn and help improve outcomes. A **co-design** approach such as this may also help develop a better understanding of the issues of consent and commercial use identified by the consultation. These issues of course can't be addressed within the bounds of theory and consultation rooms alone.

3. With regards to data sharing of deidentified personal data for policy and research, **how will re-identification risks be monitored, minimised and managed?** A growing body of research and groups such as Data61 have highlighted how de-identified personal data

can be reidentified relatively easily [6]. This must be a focus of any prototyping for data sharing involving personal data, even in aggregated or de-identified forms. Processes to minimise these risks need to be developed, and both public servants and communities need to be informed of these risks, so red lines can be identified, where the costs of the risk of re-identification outweigh the benefits of sharing personal data.

4. With regards to data sharing for policy, **how can a culture be fostered where data is shared in order to interrogate and improve policies, rather than justify them?** Data analysis is easy to get wrong. For example, there is a risk that an increased use of data sharing in government will elevate the importance of correlation without interrogating its (often weak) relationship to causation. Or the wrong analytical techniques may be applied to data. These risks could be mitigated by better data sharing with researchers. For example, where data analysis is used to assess the impact of a significant policy, we would suggest tools like (a) exposing full data analysis and methods for critique or comment; or (b) sharing data with more than one set of analysts (including researchers outside of government) to see if they reach the same result.

**TEACHING**

People in government applying these reforms need to be supported with training and mentorship, **so they can ask important questions about relevant data at appropriate times.** This does not necessarily mean acquiring technical skills; it can mean gaining a sufficient understanding to engage with *and ideally collaborate with* technical experts in confident and informed ways.

Business education, such as MBAs, may provide a useful model that could be adjusted and adapted to the context of data use in government. An MBA is a generalist degree which teaches a 'management practice'. Students learn the core practises of different technical disciplines, and learn to engage critically with these disciplines through case study learning. This sort of learning does not seek to develop experts with 'the answer'. It aims to give people in decision-making positions two things: an informed sense of what to look for, question and prioritise; and ways of engaging with *and translating for* experts in different fields. These are just some of the skills decision makers and policy people in government need when using data and working with data scientists.

**VALUES**

Lastly, as these reforms are implemented, we hope this consultation recognises that values are embedded in *how* things are done, separate from and sometimes in contradiction to values stated in theory. We hope this consultation, perhaps in coordination with the Data Commissioner, gives thought to values in the shift from theory to practise, and we offer two brief thoughts on this here:

[6] https://algorithm.data61.csiro.au/engineering-identity-from-anonymity-our-work-on-risks-of-re-identification/

**Friction can be good as well as bad.** Just as sometimes friction can create a cost, for instance by making it harder to access a service, sometimes friction can provide a benefit, for instance by protecting a human right. The value of friction changes depending on its context.

**Red lines defined in theory need to be rigorously applied in practice in order to work.** In particular, we note the red line in the Paper that data sharing is not for compliance/enforcement.

We also note two common ways that red lines around how data is used *and not used* can become blurred and so then breached:

1. where red lines are maintained in large institutions holding large amounts of data in different departments with different responsibilities (such as banks or governments),
2. where established red lines are challenged by a new urgent context (such as is often the case with sudden or new issues of national security).

These reforms must be especially alert to these issues with red lines over time. Once a new tool or process is available in one part of an institution, or accepted in one context, there can be a temptation or pressure to use it in another part of an institution or another context. Moreover, when it is within a single institution, or in the context of national security, this change can be presented as a small step, when in fact it may be a fundamental shift. This is why the red line noted above, and any others identified in this consultation, must be vigilantly maintained. Any exceptions or adjustments to red lines must be vigorously and openly interrogated. Responsibility for all this must be clearly allocated and monitored.

We believe a long term vision here is not only for Australia's governments to be better informed in the decisions they make on behalf of citizens, but for citizens to be better informed about these decisions and better able to engage with and improve them. We also believe Australia is well positioned to play a valuable role in supporting such a trend in data and sharing across open democracies and economies. Our culture tends to value the applied over the theoretical, we are willing to adopt new technologies, and we are capable of leading major systems reforms (as the world's 'democratic laboratory' in the 1890s and as far-sighted economic reformers in the 1980s). Let us not limit our ambitions and hopes here. Let these reforms embody sharing in its most holistic, hopeful and democratic sense. In this way, data sharing can indeed play a part in creating a fairer and safer society for all Australians.