THE UNIVERSITY OF SYDNEY

SYDNEY HEALTH ETHICS
SYDNEY SCHOOL OF PUBLIC HEALTH
FACULTY OF MEDICINE AND HEALTH

SUBMISSION ON DATA SHARING AND RELEASE LEGISLATIVE REFORMS

DISCUSSION PAPER AND PRIVACY IMPACT ASSESSMENT

NORAH GREWAL, RESEARCH ASSOCIATE IN LAW AND ETHICS
DR AINSLEY J. NEWSON, ASSOCIATE PROFESSOR OF BIOETHICS

16 OCTOBER 2019
Correspondence: [redacted]

We thank you for the opportunity to make this submission. We look forward to making further submissions on the proposed Data Sharing scheme as it progresses.

Our submission touches on several concepts used or implied in the Discussion Paper, namely: transparency, accountability, trust, public benefit (or interest), reciprocity and autonomy. We focus on the ‘Research and Development’ purpose and consider what a clear test for commercial uses of public sector data might look like.

The Discussion Paper says, ‘We are hoping to shift thinking about data sharing from ‘can I share?’ to ‘how can I safely share?’’.[1] The Paper’s overall tone shares its optimism with those who see ‘Big Data’ as an essential means to solve societal problems. The beneficial outcomes of data sharing are taken as a *fait accompli*, as in the following statement: ‘The public sector data reforms *must* lead to *more* and *better* data sharing...’.[2] In this light, more data simply means better statistical analyses for the current state of things and better algorithms to predict the future state of things.[3] But this overlooks concerns, including among some who study ethics, about the epistemology of ‘Big Data’.[4] We can’t discuss them in detail here, but these concerns go beyond those of privacy (identifiability), security and ‘safety’. We ask the Data Legislation Team to consider other pressing issues like bias and what is arguably only a thin veil of objectivity. Instead of, ‘How can I safely share?’, why not ask: ‘*Should* I share?’.

Trust

Many would agree that ‘building trust through transparency’ is a goal worth pursuing. But trust is a difficult concept to pin down. Trust is something the public gives (and refuses). Institutions can encourage trust, but not control or ‘build’ it. Instead, *trustworthiness* might be easier to measure and control. Transparency influences trustworthiness more than it does trust.

How can the public judge trustworthiness and place (or refuse) their trust intelligently? This too, is hard to pin down. But we can use one conceptual account as an example.[5] For the proposed scheme, the public might want to scrutinise the veracity of empirical *claims* that researchers make from the data they access. Or they might ask if Data Custodians under this scheme are *competent* in sharing data at the scale envisioned. Their focus might

[1] Department of the Prime Minister and Cabinet, *Data Sharing and Release Legislative Reforms Discussion Paper* (Discussion Paper, September 2019) (‘*Discussion Paper*’) 5.
[2] Ibid 21 (emphasis added).
[3] ‘In the age of “big data,” uncertainty is presented as an information problem that can be overcome with comprehensive data collection, statistical analysis that can identify patterns and relationships, and algorithms that can determine future outcomes by analyzing past outcomes’: Jackie Wang, *Carceral Capitalism* (Semiotext(e), 2018) 238.
[4] See, eg, Brent Daniel Mittelstadt and Luciano Floridi, ‘The Ethics of Big Data: Current and Foreseeable Issues in Biomedical Contexts’ (2016) 22(2) *Science and Engineering Ethics* 303 (‘The Ethics of Big Data’); Brent Daniel Mittelstadt and Luciano Floridi (eds), *The Ethics of Biomedical Big Data* (Springer, 2016) (‘*The Ethics of Biomedical Big Data*’); Wendy Lipworth et al, ‘Ethics and Epistemology in Big Data Research’ (2017) 14(4) *Journal of Bioethical Inquiry* 489.
[5] Onora O’Neill, *Autonomy and Trust in Bioethics* (Cambridge University Press, 2002).

instead be on the National Data Commissioner's *commitment* to holding Accredited Users accountable (to use public sector data for the 'right purposes' only). The *reliability* and *honesty* of those who make such claims or offer such commitments and competencies are also evidence of their trustworthiness. Can the public *rely* on the expertise of Data Custodians and in Australia's technical infrastructure to meet the demands of such sharing 'safely' and appropriately? This question is particularly important given that the proposed scheme 'will not provide for merits review for data sharing decisions by Data Custodians'.[6]

But as the Discussion Paper says, many Commonwealth entities already share data they hold under several other laws. We are not starting afresh. At the same time, the proposed scheme's implementation risks being clumsy because it is writ-large. We think it might be 'safer' and more appropriate to start with those entities that deal with the movement of large volumes of data as part of their remit. This might look like a **graduated approach** (i.e. using regulations and sunset clauses) that prioritises the Custodians that are most ready (and willing) to engage in data sharing.

But of course, competence is only one element that gives evidence of trustworthiness. We know, for example, that the ABS can competently handle data – but even one occasion of not doing so,[7] or the perceived dishonesty of its intentions,[8] could be enough evidence for the public *against* trustworthiness.

Transparency and Accountability

What can transparency achieve? We know that mere disclosure into the public domain is often not enough to increase trustworthiness. Judging trustworthiness relies on information that is *accessible*, *understandable* and, most importantly, *assessable*.[9] Most Australians have yet to hear about the National Data Commissioner. Few of us will know to access a 'public register of data sharing agreements' on the website of a new regulator.[10] For those who do find it and can then make sense of what they find, transparency achieves little if they cannot also judge and *act on it*.[11] The proposed *public* register does not give the public a means to challenge data sharing on its merits.

Given the emphasis on 'public benefit' throughout the Discussion Paper, there appears to be little commitment to include mechanisms that support public oversight. We ask the Data Legislation Team to give the National Data Commissioner a broad remit on 'avenues for individuals to raise issues or ask for decisions to be reviewed'.[12] This broad remit should complement mechanisms for *collective* control over the 'decisions to share, the conditions of sharing, or to deny access'.[13] Complaints of interference with individual privacy under the *Privacy Act* will not address the potential group-based harms caused by this scheme.[14] We acknowledge that the proposed scheme will no longer include sharing for compliance and assurance activities and work is being done regarding 'Indigenous data'. But the Discussion Paper still focuses more on concepts and activities that draw on a narrow and individualistic conception of autonomy (e.g. the Five Safes framework and inadvertent or intentional disclosures) rather than recognising collective, relational or group-based harms.

In saying that, placing trust in the trustworthy (and refusing trust in the untrustworthy) can be demanding on the public. Most people do not have the time to scrutinise information and even transparency done well cannot meet

[6] *Discussion Paper* 51.

[7] 'ABS Blames Overseas Hacking Attack for Census Night Shambles', *ABC News* (online at 10 August 2016) <https://www.abc.net.au/news/2016-08-10/australian-bureau-of-statistics-says-census-website-hacked/7712216>.

[8] Simon Elvery, 'Uni Hackers Identify ABS Algorithm Flaw, Fear It Puts Census Data at Risk', *ABC News* (online at 1 March 2019) <https://www.abc.net.au/news/2019-03-01/abs-census-vulnerability/10857236>; 'Truth, Damned Truth and Statistics', *Radio National* (online at 29 August 2019) <https://www.abc.net.au/radionational/programs/drive/truth-damned-truth-and-statistics/11462834>.

[9] This is also known as 'intelligent accountability': Onora O'Neill, 'Intelligent Accountability in Education' (2013) 39(1) *Oxford Review of Education* 4.; Onora O'Neill, 'Trust, Trustworthiness, and Accountability*' in Nicholas Morris and David Vines (eds), *Capital Failure* (Oxford University Press, 2014) 172 <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780198712220.001.0001/acprof-9780198712220-chapter-8>.

[10] *Discussion Paper* 36.

[11] This is not the same thing as 'digital literacy' – a nice-sounding concept that can devolve into unhelpful responsbilisation.

[12] *Discussion Paper* 51.

[13] Ibid (parentheses omitted).

[14] Mittelstadt and Floridi (n 4).

the needs of everyone. Making it easier to judge trustworthiness through transparency is not enough without adequate accountability measures. Between its 'graduated enforcement model' and its role as 'champion and advocate for greater data sharing and release' – there is little evidence of the National Data Commissioner's commitment to public benefit and intelligent transparency and accountability.[15] We understand that it can be hard to get the balance right. Some forms of accountability might impose too great a burden of compliance or create perverse incentives.[16] But at the very least, the regulator's list of 'objectives' should be updated – mere 'engage[ment] with the community' is not enough.

Commercial purposes

We appreciate that the Data Legislation Team is open to submissions that advocate for outright exclusion of all commercial purposes (as per the secondary use of My Health Record data). But this might be impractical under the proposed scheme. We agree with the Discussion Paper that a test based on sector won't work,[17] because the purpose test combines 'Research and Development' as one purpose. To exclude the private sector would also mean excluding public-private research partnerships, consortia science, research that is funded according to its 'innovation-potential' and so on.[18]

But this does not mean that the scheme should authorise sharing public sector data for *all* commercial purposes. There are clearly differences between research that involves private funding and commercial uses that are solely profit-driven. Indeed, it might be reasonable to assume the public's concerns relate to not only profit-driven motives, but also to commercialisation of public services in general, and to potential misuses of data (e.g. for surveillance and marketing). The difficultly, of course, is defining and delineating those differences to create a clear legal test for commercial uses that the public might agree with.

The paradox of privacy is that people sometimes freely give away personal data to for-profit companies. Many companies that operate in Australia hold or have access to a wealth of personal data – more so than the public sector. This gives rise to several related questions. Just how much more data do businesses need? What does public sector data offer exactly? And why should Australians feel comfortable with the state sharing it? Some submissions might say that Australians – as citizens and residents – should expect data sharing in exchange for services provided by the state. This would be a novel extension of the 'reciprocity in exchange' and 'common benefits for all' concepts.[19] Others might say that Australians already make such 'bargains' by making themselves known to receive public services or by paying taxes. For this reason, we are cautious about the Discussion Paper's repeated use of the phrases, 'innovation' and 'economic well-being'. If the criterion for acceptable commercial purposes is 'public benefit', then even solely profit-driven sharing is a 'pubic benefit' – because it might result in 'economic well-being'. It would not matter that such benefits tend to be indirect (and difficult to measure).

The point is that sharing for commercial purposes *might* result in 'economic well-being'. The outcome is not clear. And yet the Discussion Paper suggests the purpose test should be designed to 'maximise public benefits while meeting community expectations'.[20] Such a test would allow for 'commercial uses that benefit society, but do not harm individuals or businesses'. We think a consequentialist (outcomes-focused) test based on 'public benefit' will be impractical because it relies on knowing (or at least being able to quantify) the outcomes of sharing in advance. It is also difficult to make comparisons between public benefit and individual harm.[21]

[15] *Discussion Paper* 13.
[16] O'Neill (n 9).
[17] *Discussion Paper* 27.
[18] Edward S Dove and Vural Özdemir, 'What Role for Law, Human Rights, and Bioethics in an Age of Big Data, Consortia Science, and Consortia Ethics? The Importance of Trustworthiness' (2015) 4(3) *Laws* 515.
[19] Barbara J Evans, 'Much Ado about Data Ownership' (2011) 25(1) *Harvard Journal of Law & Technology* 70.
[20] *Discussion Paper* 27.
[21] This is yet another reason to consider group-based harm. 'What real balance can be struck between the vague, protean societal interest in Big Data collection and processing, be it for national security, economic well-being, or health, and the vague, protean individual harm suffered and ambiguous individual interest attached to one's data in a massive database?': Dove and Özdemir (n 18) 18.

This leaves open the opportunity for a test on commercial uses that at least indirectly allows for public oversight. Instead of focusing on ‘public benefit’ – which in this context could easily allow solely for-profit purposes – we propose a test based on the **‘reasonable expectations’** of the public that is not consequentialist and does not involve a balancing exercise. The ‘reasonable’ in ‘reasonable expectations’ would allow for such decisions to be made on a case-by-case basis, considerate of the context (i.e. ‘on the facts’) and is malleable. Importantly, such a test does not offer guidance, but neither does it set too high a threshold.

Whatever the test, the ‘Research and Development’ purpose must not obscure the fact that informational research involving personal data is usually subject to scrutiny by human research ethics committees. The Data Legislation Team should be careful to not create a discrepancy, whereby sharing for commercial purposes is subject to lesser scrutiny than research merely because such administrative mechanisms do not exist in the private sector. The National Data Commissioner’s role is vital in this regard.

Finally, the definition of ‘data’ in the Discussion Paper obscures the conceptual difference between data and information.[22] The distinction is important when thinking about ownership of and trade in personal data. The Discussion Paper says that ‘open data release drives productivity, innovation and competition’.[23] But it does not explicitly consider the potential for trade in public sector data that has been shared for commercial purposes. The public might feel strongly about this (regardless of whether it is currently possible).

[22] Václav Janeček, ‘Ownership of Personal Data in the Internet of Things’ (2018) 34(5) *Computer Law & Security Review* 1039.
[23] *Discussion Paper* 26.