Response to the Data Sharing and Release Legislative Reforms Discussion Paper

ANU Centre for Social Research and Methods

Associate Professor Nicholas Biddle[1] and Professor Matthew Gray[2]

1   Associate Director, ANU Centre for Social Research and Methods, and Member of National Data Advisory Council
2   Director, ANU Centre for Social Research and Methods

15 October 2019

1   Introduction and the need for data release and sharing

More digital information about us is being created, captured, stored and shared than ever before, and there is a greater capacity to link information across multiple sources of data and across multiple time periods. Although this creates opportunities, it also creates the risk that the data will be used in a way that is not in our best interests.

In addition to social media data and data held by commercial organisations, governments also hold a lot of information about us. When a person completes a tax return, applies for a social security benefit, completes a government survey, uses their Medicare card, or comes in or out of Australia, that information is stored and often linked to other sources of information within government. Almost any interaction with have with government has the potential to generate useful data.

Personal information held by government can be used for the benefit of individuals. For example, it can be used to make sure that an individual is receiving a benefit that they have applied for, and that payments are relatively quick and easy to receive. Data can also be used to ensure that an individual does not have to enter information more than once.

The information held by government can also be used to benefit society more broadly. Australia's current social security system is one of the most targeted in the world. What that means is we get much more benefit in terms of reducing poverty and inequality for a given dollar spent than many other comparable countries. To maintain the level of targeting we do have, we need to be able to make sure that those who aren't entitled to a program are not receiving support that could be better spent on someone else, as well as make sure those who aren't entitled to receiving benefits aren't missing out.

While data is currently being used for program design, delivery and evaluation, there is a great deal of scope for the information to be used even more effectively.

Large administrative datasets are also invaluable for research purposes. This research not only has direct benefits for policy design, but also allows us to learn about the society in which we live. Some examples of research that the ANU Centre for Social Research and Methods has been working or collaborating on are:

- An analysis of movement in and out of the very top of the income distribution using linked tax, social security, census and Medicare data (the Basic Longitudinal Extract from the Multi-Agency Data Integration Partnership)
- Measuring differences in income for vulnerable groups within society, and the extent to which these differences are geographically concentrated and hold after controlling for the rich supplementary information in these datasets;
- Analysis of the social determinants of mortality, which has shown that a person's education level is a clear predictor of their life expectancy, regardless of where they live;
- Analysis of the distribution of out-of-pocket health expenditure by socioeconomic status;

The Australian National University Centre for Social Research and Methods

- Analysis of the differences in developmental outcomes between children who have and have not participated in preschool, with a particular focus on the measured returns for Indigenous children;
- Evaluations of a number of specific government policies including Work for the Dole 2014-15; jobhelp (a prisoner pre-release labour market program) and changes to the Childcare system;
- Analysis of data from the recall of Takata airbags, to ensure that certain individuals aren't at greater risk of injury or death from faulty airbags by not replacing;
- Microsimulation modelling of the tax and social security systems and their interactions; and
- Analysis of changing Indigenous demographic and socioeconomic outcomes as part of an evaluation of the Closing the Gap targets.

All of the projects mentioned above are highly reliant on access to individual-level government data, that has been provided to the ANU through existing legislative arrangements. These research projects have had, or we expect to have significant policy impact, with the evidence potentially used to improve the lives of the most disadvantaged members of society and/or help government deliver services in ways that are more impactful or cost effective.

Access to this data has, however, been quite time intensive. Furthermore, release of some of the findings from this research to the general public and academic community has been restricted at times by those who have funded the research. It is for this reason that the ANU Centre for Social Research and Methods is highly supportive of the intent of and principles outlined in the Data Sharing and Release Legislative Reforms Discussion Paper (henceforth, 'the Discussion Paper').

Response to Question 3

Recommendation 1 – The ANU Centre for Social Research and Methods is strongly supportive of the legislative framework to achieve a more streamlined and safer data sharing

---

2   The research process

One of the characteristics of research using large administrative datasets is the iterative nature of the research process. Initial research questions and methodologies are specified before data analysis has commenced, and these can be incorporated in an application for data as part of the Data Sharing and Release Legislation. However, these research questions and methodologies adapt and evolve as the analysis progresses, including through the initial presentation of results and findings to co-authors and then peers. A number of new and innovative research questions have been identified during the data analysis phase of the projects listed above.

The Discussion Paper currently lists the following as the first two offences as a breach of legislation:

The Australian National University Centre for Social Research and Methods

- unauthorised sharing, release and use of data
- unauthorised uses of data created under Data Sharing and Release legislation

While we are very supportive of the graduated enforcement approach outlined in the Discussion Paper, in our view offences related to unauthorised sharing and release of data should be treated separately from unauthorised use of data. Furthermore, we would advocate for 'exploratory analysis' to be incorporated as a legitimate use as part of the Legislation, whilst maintaining a need to seek approval against the purposes of the research project for release of data and findings.

Response to Question 13

Recommendation 2 – Careful consideration should be given to the iterative nature of research using administrative and survey data.

---

3 Balancing of risks and potential harms

Despite the obvious benefits outlined above, there is a risk of data about us held by government or commercial organisations to cause us harm. At an individual level, data that is stolen or released without our permission can have significant consequences in terms of identity theft and extortion.

Data that is incomplete, or analysis that is incorrect has the additional potential to cause real stress and harm to individuals. Mistakes have always been made in the delivery of services. However, as more reliance is placed on automatic and data driven processing without human oversight, power imbalances could increase and the ability to overturn incorrect decisions may decrease.

Even if individuals aren't impacted directly as individuals, increasing availability and use of data can cause harm to a particular group or society as a whole. There has been increasing pressure from Indigenous peoples in Australia and overseas to have greater control over data held by governments and other organisations about them. This 'Data Sovereignty' movement recognises that a number of previous policy decisions that have had very negative effects on Indigenous peoples have been due to data held about them, and that Indigenous peoples and organisations are only able to advocate for better policies if they have access and control over their own data.

It is encouraging that the Discussion Paper recognises the specific needs of Indigenous communities and organisations within the current data environment, and we would strongly support the stated aim of continuing 'to engage and work with the community and stakeholders to ensure the right policy and practices are used.' We would recommend, however, that this not only be done through the National Indigenous Australians Agency. Rather, we would recommend that Indigenous researchers, Indigenous communities, and organisations that work closely with Indigenous peoples have the opportunity to directly

interact with the Office of the National Data Commissioner (ONDC), and that specific expertise and Indigenous representation be incorporated into the development of the ONDC.

Moreover, inherent structural biases within a system can be exacerbated by data driven decision making. A model built on the current distribution of outcomes can identify discrimination. But, it can also reinforce that discrimination. For example, Artificial Intelligence (AI) can be used by a bank to predict default on a loan, or in the criminal justice system to predict recidivism or violation of parole. If differences between groups (for example Indigenous/non-Indigenous Australians) are caused by structural or individual racism, then a person from a minority group might miss out on a loan or parole because of that AI model, even if none of their own data is used in the decision.

Response to Questions 2, 7, and 14

In our view, the potential harms outlined above can be minimised by three related policy areas:

- Recommendation 3 – Increased investment by government in the skills and capabilities of the government and research community to robustly use large administrative datasets;
- Recommendation 4 – Include as a requirement of access to government data a commitment to release the statistical code and algorithms used to generate findings; and
- Recommendation 5 – Publication by default for any research based on data shared within government or shared with researchers

---

4   Attitudes of the general public to data release and sharing

Given all the potential positives and negatives of this new data environment, it is important to monitor and respond to people's attitudes and beliefs with regards to data release and sharing.

In late 2018, the ANU Centre for Social Research and Methods, in partnership with the Social Research Centre undertook a survey of the Australian population on attitudes towards data and data governance. The results of the public opinion survey on 2,150 respondents showed that Australians are generally supportive of data being made available to researchers (especially those in universities) and being used within government.

There was much less support for multiple sources of data to be linked. This demonstrates a need to more carefully explain how such data linkage can have benefits for individuals, to set up proper safeguards for such linkage, and to not move too far ahead of public opinion.

Although the level of support for government to use and share data is generally high, there is much less support for the propositions that the current government has the right safeguards in place or can be trusted with people's data. Our findings suggest that only 34.0% of people think that the Australian Government could respond effectively to a data breach. An even

The Australian National University Centre for Social Research and Methods

smaller percentage think that the Australian Government has the ability to prevent data being hacked or leaked (29.7%); can be trusted to use data responsibly (29.3%); or is open and honest about how data are collected, used and shared (26.8%).

While there is a worryingly low level of trust in general, government and research organisations are much more trusted than commercial entities; and levels of trust in social media companies are particularly low. This evidence provides strong support for a heavy reliance on the 'safe people' aspects of the five-safes framework that is planned to be embedded in the Data Sharing and Release Legislation.

In separate analysis, the ANU Centre for Social Research and Methods have looked at what predicts people's willingness to share their data with government and to make use of the data that is currently held by government. We have analysed data related to the Consumer Data Right, opt-out of MyHealth, and consent to data linkage.

We found that trust matters. In a paper released in October 2019, we showed that low trust in government with regards to data is predictive of people not taking up the Consumer Data Right. Specifically, 42.2 per cent of those who had high trust in government with regards to maintaining the privacy of their data said that they were likely to 'Attempt to use data about you to try and obtain a better product' through the Consumer Data Right. For those who had low trust in government, this fell to 28.0 per cent.

In soon to be published analysis, we have also shown that opting out of the MyHealth system was even more highly correlated with trust in government. In October 2018, 11.1 per cent of those who had high trust in government said that they didn't want a MyHealth record and were likely to opt out. For those who had low trust in government, this rose to 37.0 per cent.

Both of these initiatives – the Consumer Data Right and MyHealth – need large take-up for the system to be viable. An individual is likely to benefit if they participate, but the system requires a large proportion of participants to increase competition (in the case of the Consumer Data Right) and to design effective health service delivery (in the case of MyHealth).

Response to Question 14

Recommendation 6 – A regular survey should be undertaken to monitor public attitudes to data sharing and release, as well as trust in government with regards to data privacy
Recommendation 7 – Major government initiatives that make use of government data should be evaluated against their effect on trust in government with regards to data privacy

---

The Australian National University Centre for Social Research and Methods

5   Further information

- Biddle, N., Edwards, B., Gray, M. and McEachern, S., 2018. Public attitudes towards data governance in Australia.
    - http://csrm.cass.anu.edu.au/sites/default/files/docs/2018/12/CSRM-WP-DATAGOVERNANCE-PUBLISH_0.pdf.
- Biddle, N., and Marasinghe, D., 2019 'Risky Data: The combined effect of framing, trust and risk preferences on the intended participation in the Consumer Data Right
    - https://taxpolicy.crawford.anu.edu.au/publication/ttpi-working-papers/15321/risky-data-combined-effect-framing-trust-and-risk-preferences

The Australian National University Centre for Social Research and Methods