THE UNIVERSITY
OF QUEENSLAND
AUSTRALIA

CREATE CHANGE

Submission in response to the Data Sharing and Release Legislative Reforms Discussion Paper

Contents

Submission authors:

Professor Mark Western, Institute for Social Science Research, The University of Queensland

Professor Mark Western is the Director of the Institute for Social Science Research and an exceptional Australian sociologist with particular strengths in social science methods. His research spans a number of areas from social inequality, to education, families, and households; with recent research experience including evaluation framework development, longitudinal data architecture and costed study designs. In recent years, Mark has advised on the design of evaluation frameworks for the Try, Test and Learn Fund and the Queensland Reform Program on Domestic and Family Violence, and partnered with The Smith Family to build an integrated dataset from their client administrative databases and bespoke surveys to evaluate their education services.

Mark has achieved particular success in developing quantitative methodologies with real-world application, such as producing the benchmark estimates of international students' expenditure that are now used in the Australian National Accounts; and leading the research teams that carried out the first national survey of information technology use in general medical practice for the Divisions of General Practice and the Department of Health. He has written chapters, articles and reports on research methodologies including quantitative methods, dual-frame sampling, survey methods and longitudinal designs, and was instrumental in the development of the case for a new Australian birth cohort study for the Australian Department of Social Services. Mark is currently a member of the DPHW Housing and Homelessness Research Alliance.

Aside from ARC Centre of Excellence funding, Mark has held research grants and government contracts worth over $20m. He has edited and authored 7 books, and over 100 book chapters, journal articles and commissioned reports. He is a Fellow of the Academy of Social Sciences in Australia (FASSA), an International Fellow of the Center for the Study of Poverty and Inequality at Stanford University, a former Chief Investigator on the ARC Centre of Excellence in Policing and Security and a former member of the ARC College of Experts.

Professor Tim Reddel, Institute for Social Science Research, The University of Queensland

Tim was recently appointed as a Professor at UQ, leading the Social Solutions Lab within ISSR. The Social Solutions Lab is aimed at creating greater public policy impact for social science knowledge and research and will draw from Tim's extensive experience across government and academia.

Tim previously led the Australian Government's Department of Social Services' Policy Office, which was responsible for evidence-based policy, research and evaluation strategies to support and enable quality strategic policy across the Department.

Tim has also held a number of senior executive roles in the Queensland public service, the community services sector and academia including leading an Australian Research Council funded project into social inclusion and place management from 2000 until 2004. He was also appointed to the Australian Research Council Engagement and Impact Assessment Panel for Social Sciences in 2018 to examine how well universities were engaging with research end-users (outside of academia) and delivering economic, social, environmental and cultural impact.

Professor Rhema Vaithianathan, Institute for Social Science Research, The University of Queensland

Rhema holds a partial appointment at the Institute for Social Science Research, University of Queensland as a Professor of Social Data and Analytics. She is also a Professor of Economics at Auckland University of Technology (New Zealand) where she is director of the Centre for Social Data Analytics (a research centre focused on using data analytics for social impact) and a Senior Research Fellow at Singapore Management University.

Her predictive analytics work focuses on the methodologies for, and implementation and implications of, predictive risk modelling in health and social services settings. She leads the international research team that developed, and continues to refine, the Allegheny Family Screening Tool, a child welfare predictive risk modelling tool for Allegheny County (PA, United States). Other predictive analytics work in the United States

includes implementation of a child welfare predictive risk model for Douglas County, Colorado and a feasibility study for a predictive risk algorithm to help Allegheny County prioritise homelessness services.

Rhema's health and human services and policy research looks at understanding and improving the way in which health and human services operate. It includes work on indigenous health and disparities. Recent New Zealand studies have explored the cumulative prevalence of child maltreatment, rates of injury and mortality among children identified as being at high risk of maltreatment, and protective factors that may allow some children to 'beat the odds' and experience no childhood adversities despite being at high risk.

Rhema is internationally recognised for leading the implementation of machine learning tools in high stakes government systems such as child welfare and is frequently invited to speak to government agencies, researchers and practitioners around the world about the ethical use of machine learning tools in public policy. She has held research positions in Australia, Singapore and the United States, including a Harkness Fellowship at Harvard University.

About the University of Queensland

For more than a century, The University of Queensland (UQ) has maintained a global reputation for delivering knowledge leadership for a better world. UQ has won more Australian Awards for University Teaching than any other university. This commitment to quality teaching empowers our 52,000 current students, studying across UQ's three campuses, to create positive change for society.

Our research has global impact, delivered by an interdisciplinary research community of more than 1500 researchers at our six faculties, eight research institutes and more than 100 research centres. The most prestigious and widely recognised rankings of world universities also consistently place UQ among the world's top universities.

About the Institute for Social Science Research

The Institute for Social Science Research (ISSR) is the flagship unit for applied social science research at UQ. ISSR is one of the largest social science research institutes in Australia based on staff and total revenue.

The Institute for Social Science Research (ISSR) undertakes solution focused research addressing societal challenges and is a national leader in advanced interdisciplinary social science and evidence-based policy research. ISSR researchers work with diverse academic disciplines and collaborate with government agencies and the private and not-for profit sectors to address key Australian and international public policy questions that span the life course and require investigation of areas such as social inequality, health and wellbeing, education, employment, housing, social services, Aboriginal and Torres Strait Islander environments research and international development. The Institute leads advances in social science and evaluation methodology, and is developing new capabilities in social data science.

Researchers at ISSR are familiar users of public sector data to inform problem analysis and changes over time for population groups due to social policy and practice implementation. ISSR's work and analysis of such data sets, often complemented by mixed methods approaches, has informed and influenced social policy and practice in Australia for more than 11 years.

 ISSR is the headquarters for the Australian Research Council Centre of Excellence for Children and Families over the Life Course (the Life Course Centre), an international collaboration of 22 organisations working to identify the causes and consequences of deep and persistent disadvantage in Australia. The Life Course Centre aims to identify the drivers of disadvantage in Australia and to develop innovative solutions to mitigate or reverse its impact.

Response to the Discussion Paper questions

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

The Discussion Paper defines data release as 'providing controlled access to the right people for the right reasons with safeguards in place' and data sharing as `open data that is made available to the world at large'. This distinction is clear.

However, we think the Discussion Paper does not attend sufficiently to relevant differences across the 3 purposes of data sharing. Section 3 describes when and how data can and should be shared for public benefit, namely; research and development, policy and programs and service delivery. The data sharing purposes may involve different actors, objectives and targets with consequently different implications for risks, safeguards, transparency and accountability mechanisms and public assurance. The data sharing purposes are also not mutually exclusive, with for instance, research and development able to be undertaken to inform the development and evaluation of government policies, programs and services and by university and non-university actors, including researchers in government. By not clearly distinguishing all three purposes, and associated factors the Discussion Paper risks creating the impression that we need a single approach to data sharing, rather than one that has a common framework, but is contextualised or nuanced by relevant differences of purpose and other things.

One potential way of clarifying distinctions among data sharing use cases involves thinking about the broad purposes of activity, the objectives of the activity, the actors involved and the targets of the activity. A potential taxonomy of data sharing is the following:

Purpose	Objectives	Actors	Targets
General Research	Contribute to general research knowledge	Researchers across sectors, including government	Research consumers, including other researchers
Policy improvement and evaluation	Development of policy evidence base Policy development and evaluation	Government policy officials and researchers (government & non-government)	Populations and cohorts subject to policy
Program and service improvement and evaluation	Development of program evidence base Program development and evaluation Program gap analysis Service delivery improvement	Policy officials, service providers, researchers, citizens	Cohorts and individuals

All three purposes potentially include research, but the other elements of the classification (objectives, actors and targets) clarify differences between Purpose types that may be relevant for the Data Sharing Principles and safeguards associated with data sharing and release (see below).

We also note that more depth and discussion on how data sharing can benefit service delivery and policy/programs would have been useful in the Discussion Paper. It seems the main service delivery focus is on service access and improving the `user experience'. There appears no real consideration as to how improved data sharing between agencies and services can support better program design that informs more tailored service delivery particularly in social and human services, e.g. risk assessments related to child protection, social investment initiatives, early intervention programs for at risk groups. Emerging technologies and techniques through AI and machine learning are front and centre in this space and are bring significant innovation to policy development, program design and service delivery but the Discussion Paper is silent on these possibilities.

Relatedly, the Discussion Paper makes cursory reference to the topic of data linkage (p. 22). In broad terms data linkage is the process that is used to combine information that relates to an individual entity from within or across multiple sources to create a richer dataset to support more targeted policies, programs, services (and research). The Discussion Paper should have at least canvassed the place data linkage has in debates around data sharing and release. Data linkage activities will necessitate greater attention to privacy impacts, public benefits and security issues, and more generally will probably complicate the implementation of safeguards built on the Data Sharing Principles.

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

We have assumed that this question relates to open data release as release to the world at large, and not controlled release of shared data. We agree that open release data needs to meet the Data Sharing Principles safeguards. We strongly believe that unit record public sector data should not be available through open data release, for instance via data.gov.au, unless there are no concerns about re-identifying the data units (for example persons, households, organisations) to which the public sector data applies. This is because the increasing availability of data held by third parties, for instance social media or private sector data, makes re-identification of public sector unit record data more feasible if third party data contains both identifying information and information that permits reliable data linkage between open public sector data and third party data. We also note that whole population unit record data removes one source of uncertainty that is an important data protection, namely whether or not a particular unit of observation is actually represented in the data. Reducing this uncertainly also increases the likelihood of re-identification.

While it might be seen as outside the legislative focus of the discussion paper, the notion of a social license to operate is critical to building and maintaining trust and transparency of controlled data sharing and release and open data release. We take this issue up again below, but we note here that a legislative approach to open release and data sharing and controlled release is not sufficient to ensure public acceptance or public trust. Trust will need to be built through the mechanisms noted in Section 5, appropriately nuanced to differences of purpose, objective, actors and targets, but trust also requires public understanding and public education. To date the strategies and tactics required for building community trust, maintaining privacy and security have not been fully articulated through the broader public data reform process. The overall social license approach should consider the differing needs and expectations of the many data stakeholders across the purposes of research and development, policy and programs and service delivery.

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

The legislative framework is a necessary, but not sufficient condition, for more streamlined and safer data sharing. There will need to be appropriate resourcing to Data Custodians and Accredited Service Providers properly implement the Data Sharing Principles, and there will also need to be education, capacity building and some culture change, within government, and among other stakeholders, including universities, service providers, privacy and "consumer" organisations. Public trust and the social licence to operate will need to be secured and actively and vigilantly monitored and maintained, because if public trust erodes, the political will to pursue data sharing and release will quickly disappear.

4   What do you think about the name, Data Sharing and Release Act?
We have no objections to it.

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

As noted in 1, the current purposes are not mutually exclusive, and do not consider other relevant factors, especially, objectives of sharing and release activity, actors, and targets of activity. We also think data linkage is a purpose of data sharing and release that is implicit in the Discussion Paper, but should be explicit and separate. Data linkage may be undertaken for one or more of the other purposes noted above, and presumably will be undertaken by Accredited Data Service Providers using data collected or generated by Data Custodians. We also think one of the examples used to explain the three purposes includes to `remind welfare recipients to report on time to get their payments' (Figure 5 page 23) could be seen to `trip' into the area of social security compliance which is seen to be outside the proposed legislation's scope.

Unpacking the purposes of data sharing also begs the question what do we mean by `data'. The Discussion Paper's description (see Discussion Paper attachment) of the term is `any facts, statistics, instructions, concepts of other information in a form capable of being communicated, analyses or processed'. It's understandable that a broad definition has been used but some explanation of the various data types would have been useful. For example, discussing the characteristics of `Big (administrative) Data' linked to the social security and taxation systems vis a vis data derived from longitudinal surveys undertaken specifically for research (whether for academic or government objectives) would have added to the depth of the discussion paper's consideration of data sharing issues.

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?

We do not think controlled release public sector data should be available for commercial uses. Our primary objection to this, as we have noted, is that such data will often be at the unit record, and many commercial organisations have additional access to other data, including their own data and that from other commercial sources, that makes re-identification of public sector data possible. This risks data breaches, creates opportunities for things like direct marketing and price and service discrimination, which may themselves be unlawful, and poses substantial risks to public trust and the social licence to operate in the sharing and release of public sector data.

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

As the Discussion Paper correctly states `there were divergent views on the balance of benefits and risks of sharing and release of public sector data'. Stakeholders such as researchers and privacy advocates often have strongly divergent views of how to balance the risks and benefits of data sharing. The data sharing principles cover a broad set of issues that could be seen as addressing the identification and management of risks such as privacy and security. In particular the threshold tests of demonstrating and specifying public benefit and addressing legal, ethical and moral considerations of a data sharing exercise are highlighted in the Principles.

We think that one way of addressing this is through a more nuanced understanding of purposes, objectives, actors and targets. To our knowledge, virtually no-one thinks that identified sensitive personal data should be available. Privacy concerns typically relate to possibilities for re-identification, inappropriate sharing and on-sharing, and other security breaches. Among "data activists" on the other hand, concerns include the fact that data sharing and release of public sector data for program and service improvement and evaluation is often focused on vulnerable marginal groups who receive government services. These are the individual targets of service delivery. Where data sharing and release is combined with clumsy or coercive service implementation, as in the recent "Robodebt" example, or with algorithmic techniques like machine learning, that are generally not transparent, or explainable, there may be particular risks that have to be managed. Particularly in the case of citizen-agency interactions for service delivery we do not think the Data Sharing Principles are strong enough in giving citizens, who will receive services, sufficient transparency, accountability, or recourse. In this case, citizens need to be able to access the data about them that is used for service delivery, they need a transparent understanding of how algorithms and data are used, they need the ability to easily correct data errors, and they need appropriate recourse if they are disadvantaged through the incorrect use of shared data.

We also note that acceptable risk standards in the Data Sharing principles will vary by purpose, and objective. For example, data sharing for academic research may be able to accept a linkage rate of 85-90% (i.e. around a 10% tolerance for mismatch)- but this is simply not acceptable for service use (where in some cases such a criminal justice history, you need a match rate close to 100%). Entity resolution - especially where there is not a single citizen ID is a much more critical issue when sharing data for services, than when sharing data for general research.

8   Is the Best Practice Guide to Applying Data Sharing Principles helpful? Are there areas where the guidance could be improved?
No comments on this.

9   Do the safeguards address key privacy risks?

See our comments above. We would like to see a more thorough working through of the Data Sharing Principles that captures the complexity of different purposes, objectives, actors and targets. We do not think unit record public sector data should be made available for open release or commercial use, unless re-identification is not a concern.

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

We think that the core principles as described in 6.3 are adequate, but the devil will be in the detail. In general terms we note that the governance criterion will need more development, particularly in relation to the use of data for services. We have already highlighted this in the case of data sharing for services where technical demands on data quality, in relation to things like match rates and entity resolution may need to be set very high, before use of linked data is considered acceptable. Governance frameworks for data sharing for services may also need to incorporate a much stronger role for citizens than in other use cases, as well as higher standards of transparency and accountability to citizens. Accreditation criteria also need to recognise that data science methods for data management and use will rapidly become much more widely available. The skills and capabilities needed to use these new methods appropriately with shared public sector data go well beyond the technical skills and expertise associated with disciplines like mathematics, statistics and computer science. The accreditation principles need to recognise that skills and capabilities to protect, management and appropriately use shared public sector data are at least as important as technical skills and expertise. Likewise, privacy standards and governance frameworks need to be robust to the new possibilities (for public good and social harm) occasioned by "big shared public sector data", new computational methods, and advances in computing power.

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 discussion paper's section 5 – Building Trust Through Transparency highlights the significant issues of trust, public expectations and community confidence in the successful management of shared public sector data release. As we have already noted we do not think the Discussion Paper sufficiently acknowledges how transparency and accountability mechanisms will themselves need to be fit for different purposes and objectives. We also note that the nuances of social licence (which we understand as active endorsement of public sector data sharing and release for specified purposes), public trust, and accountability are probably not explored in sufficient detail across the different purposes. In particular, a legislative framework will not be sufficient to secure and maintain the social licence.

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

We are most concerned about the complaints and review rights available to citizens who are subject to decisions made with shared public sector data. There is not enough information in the Discussion Paper for us to determine if the combination of existing avenues and merits review and judicial review will go far enough.

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

No comment.

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

The objectives and functions of the National Data Commissioner including the ongoing formal guidance role through legislative instruments such as Data Codes and non-binding approaches such as sharing best practice are welcomed. It will be important, however, that these approaches take into account the differing

needs of stakeholders and the variability of Australia's data environment (federated system of government differing data categories and data sharing purposes etc). The Commissioner also has a public facing role and could be seen as a `data champion' in contributing to public debates and supporting the social license to operate. We would like to see a considered and appropriately resourced approach to this issue, which we feel is not given sufficient attention at the moment. The social licence to operate complements the legislation in establishing the conditions which allow support public data sharing and release. More broadly the overall institutional architecture of the data reforms underpinned by the proposed legislation appears sound. In addition to the Commissioner's leadership role Discussion Paper also considers the critical role of Data Custodians across the data system in the operationalisation of sharing and release arrangements.